M. TECH. THESIS

On

Characterizing Noise in IBM-Q Devices using Unitarity Randomized Benchmarking

Author

Adarsh Chandrashekar

M.TECH in COMPUTER SCIENCE

INDIAN STATISTICAL INSTITUTE, KOLKATA

July 2020

Characterizing Noise in IBM-Q Devices using Unitarity Randomized Benchmarking

A PROJECT REPORT

*Submitted in partial fulfillment of the
requirements for the award of the degree*

of
MASTER OF TECHNOLOGY
in

COMPUTER SCIENCE

Submitted by:
Adarsh Chandrashekar
CS1827

Guided by:
Dr. Goutam Paul
*R.C Bose Centre for Cryptology,
Indian Statistical Institute,
Kolkata, 700108 India*

CANDIDATE'S DECLARATION

I hereby declare that the project entitled **Characterizing Noise in IBM-Q Devices using Unitarity Randomized Benchmarking** submitted in partial fulfillment for the award of the degree of Master of Technology in Computer Science completed under the supervision of Dr. Goutam Paul, ISI Kolkata is an authentic work.

Further, I declare that I have not submitted this work for the award of any other degree elsewhere.

Signature and name of the student with date

CERTIFICATE by Supervisor(s)

It is certified that the above statement made by the student is correct to the best of my knowledge.

Signature of Supervisor(s) with dates and their designation

Abstract

Quantum computers have potential applications in cryptography, optimization problems and recently proposed quantum-enhanced machine learning. Theoretical research into this novel framework of computation has found ways to complement and enhance the abilities of classical computers. Practical large scale quantum computers with sufficient computing capacity suffer from various kinds of errors, which destroy any meaningful computation.

Fault tolerant computation is an active area of research to realize quantum computers that are robust to noise. Quantum threshold theorem states that, if the physical error rates are below certain threshold (estimates give a figure around 1%) then by using error correction schemes, one can push the logical error rates to arbitrary low values. Benchmarking or quantifying the effect of noise induced in practically implementing quantum gates has been a topic of interest in recent years. Several protocols to effectively benchmark the physical error rates have been proposed that empirically calculate various metrics to quantify error rates of quantum gate operation irrespective of how the quantum computer works. This thesis is about one such protocol known as Unitarity Randomized Benchmarking (URB).

In this paper, we have *firstly*, implemented the URB protocol using Qiskit which is an open source library to program and execute quantum programs in simulators or on actual IBM-Q devices. To the best of our knowledge, our work presents the first step in the direction of experimentally benchmarking the 'coherence' of noise induced during execution of quantum programs on any commercially available device. *Secondly*, we experimentally validate the effectiveness of our slightly modified URB protocol, to gauge the 'coherence' of noise channels. We perform URB experiments by simulating two well studied noise channels namely, the single qubit depolarising channel and the single qubit bit-flip channel and as a result, validate our implementation by comparing with the theoretical unitarity values. *Thirdly*, we propose yet another modification to the URB protocol, which we call the Native Gate URB, in order to study the noise in the native gates, i.e., the gates ultimately into which the quantum circuits are compiled and executed in the actual IBM-Q devices. *Finally*, We comment on the ability of Native Gate URB protocol to detect cross talk. We find experimentally and to the best of our knowledge, that both the IBM-Q processors have noticeable but considerably low cross talk when applying native CNOT gate, based on the unitarity values. However, there is considerable difference ($\sim 10^{-2}$) in the unitarity of CNOT gate for 5-qubit processor (burlington) as compared to 15-qubit one (melbourne).

Keywords: Quantum Computing, Randomized Benchmarking, Fault tolerant Computation, IBM Quantum Experience, IBM Qiskit

Contents

List of Tables

List of Figures

4

Chapter 1

Introduction

1.1 Hello Many Worlds

1.1.1 Motivation

Quantum Computing is novel framework, to perform computation which is fundamentally different from how information is processed in classical machines. The striking difference between quantum computing principles and that of classical computing, is the ability to dramatically increase the scale of computation and perform non trivial information processing tasks which are guided by peculiarities or 'spookiness' of quantum mechanics. There are many motivating reasons (other than purely complexity theoretic reasons) to study this framework and think of it's practical implementation and application in real world use cases.

In the recent years, the computational requirements have sky rocketed with the advent of Deep learning and processing of Big Data, such as the data accumulated from social media which are highly unstructured and the sheer size of it is in the order of petabytes. This has resulted in the manufacturing of high throughput processing units that efficiently perform certain mathematical operations, mostly linear algebra routines, for instance, GPU (Graphical Processing Units) or TPU (Tensor Processing Units). Simply the need of storing this huge data in data warehouses is increasing at a striking rate and is bound to increase even more in the future.

The computational power increases with the increase in the number of transistors in the processor or core. But through the technique of VLSI (Very Large Scale Integration) used to manufacture these processors must then decrease the size of the transistors present in the core. The famous Moore's law [1] which states that the computational resources are being consumed at a rate such that the number of transistors in a core is doubling about every two years. So, the transistor sizes are reducing by a factor of two every about two years. Continuing this trend dictated by the Moore's

law, ultimately the size of a single transistor must decrease to the size of a single atom, from the current size of a few hundreds of atoms large. When that happens, due to quantum tunnelling, it will be hard to maintain fault tolerant computation in classical computers as it will make the charge present in a DRAM transistor itself uncertain, so even a single bit of information will become uncertain.

Quantum computing offers an alternative solution, by defining new ways to store information as quantum data and allows one to perform truly parallel computation thereby increasing the computational power. In the field of cryptography, the famous Shor's Algorithm [2] gave the first significant evidence that quantum computers may perform some tasks (here, is was factoring large primes) more efficiently than classical computers. Several quantum algorithms were then proposed, which were toy problems that theoretically showed that quantum algorithms are certainly better than their classical counterparts [3, See Deutsch-Jozsa, Simon's Algorithm, Bernstein-Vazirani Algorithm] which is termed as the 'quantum advantage'. On the other hand, in 2019, Google claimed to experimentally demonstrate the 'quantum supremacy' on a sampling problem, comparing the performance of Sycamore to IBM's Summit supercomputer [4]. 'Quantum supremacy' is said to have been demonstrated when a problem is easily solved by a quantum computer, while a classical computer takes unreasonable amount to solve it, i.e. extreme quantum advantage so to speak. One can say that from a theoretical perspective, this was important since, quantum computers became a serious candidate to challenge the Extended Church-Turing Thesis [5] but at the same time, it also fuelled the research into real world applications, that are significant and not efficient to solve on classical machines.

Quantum computing today, has evolved from being a theoretical framework to perform computation that mimics nature, to a practical means of computation for both industrial and academic research purposes and with promising potential applications in a variety of fields ranging from machine learning [6, 7, 8, 9, 10] and optimization problems to quantum chemistry problems for drug manufacturing to many body quantum simulations [11, 12, 13]. In the recent times, big technology companies and research labs around the world have developed indigenous quantum archetypes for research and industrial purposes which are based on different technologies, from superconducting transmon qubits (such as IBM-Q devices or Google's Sycamore)[14, 4], to ion-trap based (such as the one manufactured by Ion-Q[15]) to photonic devices [16]. Today's quantum computers fall under the category of NISQ (Noisy-Intermediate Scale Quantum) [17] devices,which have few qubits (around 10-50 qubits) and are noisy in the sense, that error detection and correction schemes are required to perform fault tolerant computation. Significant research is being done into scale these computers to the point, they can demonstrate practical quantum advantage in real world applications.

1.1.2   Postulates of Quantum Mechanics

Below, we state the basic postulates of quantum mechanics which will be the underlying principles behind the framework of quantum computing. We define these postulates in the way which is defined in the description of quantum mechanics, historically termed as the 'matrix mechanics' (which of talk of quantum systems as states in a vector space) different from 'wave mechanics' (which talk of wave functions of quantum systems), although both descriptions are equivalent in some sense. Note that, these postulates talk of pure quantum systems (which we shall define later) and the postulates are slightly changed when we want to talk of general quantum systems (or ensembles).

Postulate 1 *States of a quantum system are represented by vectors of unit length in a Hilbert space (i.e. complex vector space with inner product).*

States are usually written in Dirac-ket notation as $|\psi\rangle$, which are vectors associated with the Hilbert space $\mathcal{H}$. Associated with the Hilbert space $\mathcal{H}$ is the dual vector space $\bar{\mathcal{H}}$ such that the (natural) inner product between two vectors in $\mathcal{H}$ say, $|\psi\rangle$ and $|\phi\rangle$ is given by, $|\psi\rangle$ multiplied with dual vector of $|\phi\rangle$ written in Dirac-bra notation as $\langle\phi|$. Succinctly, the inner product is written as

$$\langle\phi|\psi\rangle = \langle\psi|\phi\rangle^*$$

where, $*$ denotes the complex conjugate. The vectors in $\mathcal{H}$ and $\bar{\mathcal{H}}$ are related by,

$$|\psi\rangle = \langle\psi|^\dagger$$

where, $\dagger$ represents the Hermitian conjugate or adjoint of an operator (i.e. in terms of matrices, it is the combined operation of transposition and taking complex conjugate of all entries in the matrix). Valid quantum states are represented by unit length vectors in this space i.e,

$$\langle\psi|\psi\rangle = 1.$$

In particular, from this postulate, we get the idea of 'superposition' of states as a consequence. For any two states $|\psi\rangle$ and $|\phi\rangle$ and two scalars α and β in $\mathbb{C}$ i.e. field of complex numbers, we get a new state $|\xi\rangle$ as,

$$|\xi\rangle = \alpha\,|\psi\rangle + \beta\,|\phi\rangle.$$

Note that, mathematically speaking, the vectors representing valid quantum states do not form a strict vector subspace of $\mathcal{H}$, although one can transform any vector in $\mathcal{H}$ to a valid state by normalizing it i.e dividing the vector by it's norm with respect to the inner product in $\mathcal{H}$ i.e dividing a vector $|\psi\rangle$ by $\langle\psi|\psi\rangle$. This will be important, when we discuss *measurement principle* in quantum mechanics.

Postulate 2 *Physical properties associated with a quantum system, such as spin, energy, momentum etc are represented by Hermitian operators (also known as self-adjoint operators) acting on the Hilbert space $\mathcal{H}$. They are also known as 'observables'.*

A Hermitian operator $\mathbf{A}$ is defined to be it's own Hermitian conjugate i.e.

$$\mathbf{A} = \mathbf{A}^{\dagger}.$$

These operators change the quantum state by acting on it, which mathematically is represented by left multiplication of matrix associated with $\mathbf{A}$ to the vector $|\psi\rangle$

$$\mathbf{A} : |\psi\rangle \rightarrow \mathbf{A}|\psi\rangle$$

where, left side $\mathbf{A}$ should be treated as abstract operator while right side $\mathbf{A}$ is the matrix representation of operator $\mathbf{A}$. We will use the same notation of an observable to mean both it's matrix representation and the abstract operator.

Observables represent the information that we can extract from a quantum system, which can be any valid physical quantity. Note here, that this 'extraction' or retrieving some physical information about the system is itself mathematically described by action of an operator. All operations (interactions) that we can perform on the system are represented by operators acting on the corresponding Hilbert space. They may or may not result in valid quantum states with normalization required if later is the case. Observables are also important in viewing a quantum system with respect to certain physical quantity of interest, so that we now talk of evolutions and interactions of the systems with respect to say their spin interactions, position etc, as described more clearly in the next postulate.

Postulate 3 (Measurement Principle) *: Measurement is an interaction with a quantum system, which measures some physical property i.e observable of the system. Following are the properties of this interaction:*

- *The only allowed values that can be observed are the eigenvalues a_n for $n = 0, 1, 2..., m$ (where, m can be smaller or equal to dimension of the Hilbert space $\mathcal{H}$) of the corresponding observable $\boldsymbol{A}$.*

- *(Born's Rule) The probability of obtaining a particular eigenvalue a_n when the measurement is performed, is given by the squared absolute value of the inner product of the state of the quantum system say, $|\psi\rangle$ with the eigenstate $|a_n\rangle$ corresponding to the eigenvalue a_n i.e,*

$$\boldsymbol{Pr}(a_n \text{ is the measurement outcome}) = |\langle a_n|\psi\rangle|^2$$

- *(Collapse of the 'wavefunction') After the quantum system is observed the state of the system becomes the normalized eigenstate $|a_n\rangle$ of the measured outcome of the experiment i.e a_n.*

9

This postulate gives the subject it's probabilistic interpretation and 'quantum' (discrete) nature. The state description of the system enocodes both the probabilistic nature (Born Rule) and 'quantum' nature (measurement outcomes belong to a discrete set) which gives the subject it's notorious difficulty in providing physical interpretation.

One can show that, Hermitian operators have real eigenvalues. If a_n is an eigenvalue of Hermitian operator $\mathbf{A}$, corresponding to the eigenstate $|a_n\rangle$ then,

$$\mathbf{A}\,|\psi\rangle = a_n\,|\psi\rangle$$

Taking adjoint both sides; adjoint of a scalar is just the complex conjugate,

$$\implies \langle\psi|\,\mathbf{A}^\dagger = a_n^*\,\langle\psi|$$

Since, $\mathbf{A}$ is Hermitian.

$$\implies \langle\psi|\,\mathbf{A} = a_n^*\,\langle\psi|$$

Left multiplying by $\langle\psi|$ in the first equation and right multiplying by $|\psi\rangle$ in the third equation,

$$\implies \langle\psi|\,\mathbf{A}\,|\psi\rangle = a_n^*\,\langle\psi|\psi\rangle = a_n\,\langle\psi|\psi\rangle$$

$$\implies a_n^* = a_n$$

i.e, a_n is real.

Also, one can show (along the same lines as above), if the vector space is equipped with an inner product, then eigenstates corresponding to different eigenvalues are always orthogonal with respect to the equipped inner product. In other words, one can show that if

$$\mathbf{A}\,|\psi\rangle = a_n\,|\psi\rangle$$

and

$$\mathbf{A}\,|\phi\rangle = a_m\,|\phi\rangle$$

for two eigenstates $|\psi\rangle$ and $|\phi\rangle$ with different eigenvalues a_n and a_m, $n \neq m$ then,

$$\langle\psi|\phi\rangle = \langle\phi|\psi\rangle = 0.$$

If all the eigenvalues are different (non-degenerate) then, the eigenstates of the observable $\mathbf{A}$ form an orthonormal basis (if we normalize the eigenstates) for the Hilbert space associated with the system $\mathcal{H}$. This means that any valid quantum state say, $|\psi\rangle$ can be written in a superposition of eigenstates of an observable $\mathbf{A}$ (since, it is the definition of a basis of a vector space).

$$|\psi\rangle = \sum_n \alpha_{a_n}\,|a_n\rangle$$

where, α_{a_n} is the complex scalar coefficient multiplying the eigenstate $|a_n\rangle$ associated with eigenvalue a_n.

We can give more meaning to the complex scalar coefficients stated above. Technically, they are called the *probability amplitudes* because of the Born's Rule. The Born's Rule gives the probabilistic interpretation to the state description of the system with respect to a basis associated with the eigenstates of an observable. Using the Born Rule we get,

$$\mathbf{Pr}(a_m \text{ is the measurement outcome}) = |\langle a_m|\psi\rangle|^2 = \sum_n |\alpha_{a_n} \langle a_m|a_n\rangle|^2$$

where, we are left multiplying by $\langle a_m|$ on both sides. Now, using the result derived above, eigenstates of different eigenvalues are orthogonal, the sum boils down to,

$$\mathbf{Pr}(a_m \text{ is the measurement outcome}) = |\langle a_m|\psi\rangle|^2 = |\alpha_{a_m}|^2.$$

Therefore, $|\alpha_{a_m}|^2$ associated with eigenstate $|a_m\rangle$ give the probability that the outcome of the measurement will be the eigenvalue a_m. By following this interpretation, we can physically give meaning to the orthogonality of the eigenstates of an observable. Physically, this represents the fact that, different outcomes of an observable correspond to physically distinguishable states. This is a trivial, but key feature that gives quantum mechanics a reproducible nature, in spite of being probabilistic, so that the results of the theory can be experimentally verified. If we prepare a state in the eigenstate of an observable, then we will see that whenever we measure the value of observable $\mathbf{A}$, we will always get the result as the corresponding eigenvalue and not any other.

The expected value of an observable $\mathbf{A}$ in some arbitrary state $|\psi\rangle$ expressed with the basis given by the eigenstates of $\mathbf{A}$ is,

$$\langle\mathbf{A}\rangle = \langle\psi|\,\mathbf{A}\,|\psi\rangle = \sum_n a_n|\alpha_{a_n}|^2 \tag{1.1}$$

One can derive above equation by using previous equations. This makes intuitive sense when we compare to the usual definition of an expectation in statistics. Note here, that in order experimentally calculate the expected value of an observable, we have to take several *counts* i.e we must have copies of the same arbitrary state, measure each one of them, record the measurement outcome and then take the weighted average of all the outcomes. One can also deduce from this, that outcome of the measurement, when the system is in an eigenstate of an observable is fixed and not probabilistic.

The collapse of the wavefunction represents, what happens to the system when measuring some observable. This interpretation is also known as the Copenhagen Interpretation of Quantum Mechanics [3, See Postulates of Quantum Mechanics] which is the widely accepted theory of measurement of quantum systems. Measurement or observation has always been a subject of active debate even today, regarding the mechanism behind which the collapse of the wave function or the state description

takes place. Another popular theory proposed by Hugh Everett was the Many-Worlds Interpretation [18] which states that the wavefunction does not collapse but, there exist some other universe where the wavefunction collapsed to all other possible values upon measurement.

We can mathematically represent the measurement process similar to an operator acting on the state in an Hilbert space. The one we use in this thesis throughout is the POVM (Positive Operator Valued Measurement).

Definition 1 (POVM Measurement) *: A POVM is a set of positive semi-definite matrices F_i on a Hilbert space $\mathcal{H}$ that sum to the identity matrix $\mathbb{I}$, i.e.*

$$\sum_{i=1}^{n} F_i = \mathbb{I}$$

. When a quantum system is observed, then a measurement outcome of i corresponding to the positive semi-definite (eigenvalues are non-negative) F_i is mathematically expressed as the action of F_i on $|\psi\rangle$. The state of the system after measurement is given by $\frac{F_i|\psi\rangle}{\langle\psi|F_i^\dagger F_i|\psi\rangle}$. One can check that this corresponds to the eigenstate associated with the outcome i. The operators F_i are usually projection operators onto the eigenspace corresponding to the outcome i.

Postulate 4 *The time evolution of quantum systems are governed by unitary evolution.*

Here, we talk about another type of interaction with the system, which is different from measurement and does not lead to wave function collapse.

Unitary operator U is a linear operator acting on the Hilbert space $\mathcal{H}$ such that

$$UU^\dagger = U^\dagger U = \mathbb{I}.$$

Why unitary operators ? Since, time evolution in quantum mechanics is reversible (makes no difference between forward and backward evolution of a system) given no measurements are being made, we naturally lead to the conclusion that if we have two orthogonal states, say two eigenstates corresponding to two different eigenvalues, then under this transformation, they must evolve again into two orthogonal states. This preserves the history of the states. Indeed for every unitary operator U there exists $U^\dagger$ such that the later 'undo' the computation of the former. Not only orthogonal states but unitary operators preserve the inner product between any two given quantum states, before and after measurement (which is again not hard to prove). It can also be shown that, corresponding to every observable $\mathbf{A}$ there exists a unitary operator $e^{i\mathbf{A}t}$ and vice-versa, where t is a scalar (can be treated as time elapsed).

Unitary evolution plays an important role in quantum computation. But, not all interactions with a quantum system are unitary for instance, measurement. Apart from measurement, there are general forms of interactions that we discuss below.

1.1.3   Density matrix formalism and quantum ensembles

It was found that, not all quantum systems can have a state vector representation. The problem arises when we want to have a mathematical description of a subsystem which is part of a larger *entangled system*. Consider the maximally entangled Bell state known as the triplet state of two qubits (in this case, two electrons with their spin observable)

$$|\psi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$$

. This state has the property that, if we measure the spin observable, (which for each qubit (electron) can be either 1 or -1 corresponding to eigenstates $|0\rangle$ or $|1\rangle$) even if one of the qubits is measured, we can instantly know the state of the other qubit, since the state when collpased, can collapse into either state with both spins down i.e $|00\rangle$ or both spins up i.e $|11\rangle$. *Entanglement* in quantum mechanics refer to this correlation between two quantum systems (i.e particles such as electrons or whole atoms etc). We will define entanglement more formally below. One can show that, there exists no state vector in the Hilbert space associated with a single spin, which can be used to describe either of the spins. This is counter intuitive in the sense that one can only know about the system when taken together, but cannot formally describe each qubit (each subsystem) as a quantum state in it's own right (within the postulates describes in the previous section).

Another motivation behind the density matrix formalism is to mathematically describe systems known as *quantum ensembles* or *quantum mixtures*. These are statistical mixtures of quantum systems, which represent our lack of knowledge about the system we are studying. For instance, a mixture of 1000 qubits, 500 in state $|0\rangle$ and rest in state $|1\rangle$ does not have a state vector representation and we have to invoke density matrices into the picture. Note that, the density matrix formalism 'contains' the postulates described previously and is a general formalism of quantum mechanics. Consider a single quantum system represented with respect to a Hilbert space $\mathcal{H}$.

Definition 2 (Density Matrix) *: A density matrix ρ is a positive semi-definite linear operator acting on the Hilbert space $\mathcal{H}$ which has the following properties:*

- *$\rho = \rho^\dagger$ (Hermiticity)*

- *$Tr(\rho) = 1$ (Normalization)*

- *$\rho \geq 0$ (Same as saying positive semi-definiteness)*

The matrix form of these density matrices are written with respect to some basis of the space $\mathcal{H}$. One can show that, using the above properties, that the set of density matrices form a convex set (proof omitted).

Theorem 1 *The set of density matrices acting on $\mathcal{H}$ form a convex set $\mathcal{D}$ i.e for two density matrices ρ and σ in $\mathcal{D}$ with $\lambda \in [0,1]$*

$$\lambda\rho + (1 - \lambda)\sigma \in \mathcal{D}$$

The extreme elements of this convex set, i.e. the density matrices which cannot be written as a convex combination of two other density matrices are exactly the *projection operators* (square of a projection operator is the operator itself), which are used in POVM. One can also prove the fact that all projection operators are just the outer product of some vector in $\mathcal{H}$ with itself. These are the *pure state density matrices*, i.e that quantum states which have a state vector representation. By construction, one can see that every state vector in $\mathcal{H}$ can be associated (mapped) to a projection operator, which is the pure state density matrix representation.

Definition 3 (Pure state) *: A pure state is of the form $|\psi\rangle\langle\psi|$ where, $|\psi\rangle$ is a vector (normalized) in the Hilbert space $\mathcal{H}$.*

Now, we can try to put things together and define general density matrices which are the most general descriptions of quantum states.

Definition 4 (General Density Matrix) *: From the theorem stated above and the definition of pure states, we can conclude that a general density matrix ρ can be written as*

$$\rho = \sum_k p_k |\psi_k\rangle\langle\psi_k|$$

For some real numbers $p_k \in [0,1]$ such that $\sum_k p_k = 1$.

The general density matrices can then be thought of statistical mixtures of pure states, which represents our lack of knowledge about the exact state of the system. Sometimes, (for instance, in entangled states) the description about the quantum subsystem can be stated as a statistical ensemble of pure states, and the coefficients (p_k in the above definition) are the probabilities with which we 'think' that the quantum state when observed will collapse to the respective pure states.
All density matrices which are not projection operators, are representations of *mixed states*. We can also write this more formally as follows.

Definition 5 (Mixed State) *: If a density matrix ρ, is such that $Tr(\rho^2) < 1$, then the density matrix represents a mixed state. One can also show that, $Tr(\rho^2) = 1$ iff ρ represents a pure state, i.e ρ is a projection operator.*

Till now, we looked at general state representations for a single quantum system. When we want to talk of multiple quantum subsystems together as part of large system, we consider the question, whether we can describe each subsystem irrespective of the other subsystems or not. This kind of independence associated with the subsystems is termed as *separability* and we call the larger system as *separable*. If a quantum state with possibly many subsystems, is separable, then we can safely ignore the subsystems that are not of our interest since, the subsystems are 'non-interacting' and we then treat our subsystem of interest different from the larger system, as a standalone. Similar to the notion of mixed states being the most general descriptions of

a system with a single Hilbert space $\mathcal{H}$, we can talk of general state description for say, a combined large system of two quantum systems with respective Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ and the Hilbert space for the combined system is then represented by the direct product of the two spaces i.e $\mathcal{H}_A \otimes \mathcal{H}_B$. We formally define the notion of a separable quantum system (with two subsystems as mentioned previously)

Definition 6 (Separable State) : *A quantum system represented with density matrix ρ_{AB} acting on the Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ (which is the direct product of the Hilbert spaces associated with subsystems A and B) is said to be separable, if ρ_{AB} can be expressed as*

$$\rho_{AB} = \sum_i p_i \rho_{A_i} \otimes \rho_{B_i}$$

where, ρ_{A_i} and ρ_{B_i} are density matrices acting on the Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively and $\sum_i p_i = 1$, $p_i \in [0, 1]$.

This definition can be easily extended for systems having more than two subsystems. Each of the terms $\rho_{A_i} \otimes \rho_{B_i}$ are called *product states*, which means that we have a description of each subsystem independently of the other (in other words, they are uncorrelated). Essentially, a separable quantum system is a probability distribution over the possible product states of it's subsystems. Note that, here too we can have the knowledge of a subsystem if we know the state of the larger system, and measure other subsystem, similar to the Bell state example discussed previously. But we said that the Bell state is an example of entangled system. Entanglement comes into picture, when the assumption of separability fails and we cannot write in the form mentioned in the above definition as a probability distribution over product states. This means that, the subsystems are interacting in such a way, that we cannot know the complete information about a subsystem by ignoring the other subsystem. However, we can represent our limited knowledge as a density matrix, called the *reduced density matrix* by the operation of *reduce trace* which mathematically represent the idea of 'ignoring' a subsystem.

Definition 7 *The information (possibly limited) about a quantum subsystem extracted by 'ignoring' the other subsystem is mathematically represented by the reduce trace operation. Suppose we ignore subsystem B and want to know the information about subsystem A, $\bar{\rho}_A$ from the combined density matrix ρ_{AB}. We perform the reduce trace operation as follows*

$$\bar{\rho}_A = Tr_B(\rho_{AB})$$

Suppose, $P \in \mathcal{H}_A \otimes \mathcal{H}_A$ and $Q \in \mathcal{H}_B \otimes \mathcal{H}_B$ are some linear operators over respective Hilbert spaces then, $Tr_B(P \otimes Q) = Tr(Q)P$ i.e tracing over the basis for the Hilbert space to be 'ignored'.

Note that, for a separable state ρ_{AB}, the mixed state representation we get after 'ignoring' subsystem B, is exactly the same if we had multiple copies of the ρ_{AB} and

measured both the subsystems and formed a distribution over the states that we got for subsystem A. This is the essence of a separable state, that we can ignore the other subsystem and still extract maximum information about the other subsystem. In entangled systems, the reduce trace operation provides us with less information about the state of the subsystem than what we could know, when we have both the subsystems. In particular, the mixed state description is different, if we take reduce trace or not. One can show that, the density matrix of the triplet state mentioned previously is a non-separable state and hence is entangled. But the general problem of deciding whether a system is entangled or not for arbitrary number of subsystems is NP-hard [19].

Definition 8 (Entangled State) *: (For a system made of two subsystems) A state ρ_{AB} composed of subsystems A and B is said to be entangled, if it is not separable.*

1.1.4 Introduction to Quantum Computing

We've given a brief introduction to Quantum Mechanics and defined the tools and machinery to understand quantum computing. For further clarification, we refer the reader to [3] for more on quantum information theory, quantum computation and related fields. Now, we briefly review the basic elements of quantum computing and different phases of a typical quantum computation.

Qubits

The basic unit of information in quantum computing is the idea of *qubit*. Formally, it can be any quantum system (electron, photon, ions etc) and associated with it some observable such as spin, energy etc limited to only two levels i.e, the Hilbert space of a qubit is two dimensional. In other words, when the measurement is made on a single qubit, there are only two possible values that can be obtained. We can also think of performing computation using quantum systems associated with d-dimensional Hilbert space, called *qudits*. Computation with qudits is also an area of active research, although equivalence of both the models with respect to computational power is debatable.

The quintessential example of a qubit is described by the spin of an electron in say Z direction and the associated observable is written as σ_z. The basis for the Hilbert space of the qubit is given by $\{|0\rangle, |1\rangle\}$, where $|0\rangle$ is usually the 'up' spin or the eigenstate corresponding to eigenvalue of 1, of the observable σ_z and $|1\rangle$ is the 'down' spin or the eigenstate corresponding to eigenvalue of -1 of σ_z. By the postulates of quantum mechanics, a general qubit state is given by

$$|\psi\rangle = \alpha |0\rangle + \beta |1\rangle$$

where, $\alpha, \beta \in \mathbb{C}$ and $|\langle\psi|\psi\rangle|^2 = 1$. Other observables can be the spin along X (σ_x), Y (σ_y) or any arbitrary direction. The matrices corresponding to the spin observables along X,Y and Z directions together with the 2×2 identity matrix form the *Pauli group* for two dimensional Hilbert space and the matrices $\{\sigma_x, \sigma_y, \sigma_z\}$ are known as the *Pauli matrices*.

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Here, the basis chosen to express the Pauli matrices is the $\{|0\rangle, |1\rangle\}$, sometimes also known as the *computational basis* or the *default basis*. This is the basis which is used most widely in quantum computation. At any point of computation, we are allowed to change our basis from one observable to other, which may be useful in some scenarios. For example, with respect to observable σ_x the basis is given by $\{|+\rangle, |-\rangle\}$ where, $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ has eigenvalue 1 and $|-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$ has eigenvalue -1, written in the computational basis.

Other examples for a qubit may be the the ground state and the first excited state acting as the two levels of the qubit, which is used in trapped ions archetype [20]. In superconducting transmon or Xmon qubits [21] the two levels are designated to the ground state and excited states of 'artificial atoms' which are Cooper-pairs of electrons. In terms of qubit quality, it has been observed that Trapped Ions have longer decoherence [22] times as compared to superconducting qubits although the gate times of later are much shorter than the former [23].

Quantum Register

Similar to the terminology in classical computer architectures, where an array of bits is called a bit register, an array of quantum bits or qubits is called a *quantum register*. Mathematically, a n-qubit quantum register is a quantum system, represented by the direct product ($\otimes$) of the two dimensional Hilbert space with itself n times. Since, the computational basis for a single qubit is the set $\{|0\rangle, |1\rangle\}$, naturally the computational basis for the n-qubit quantum register is the set $\bigotimes_{i=1}^{n} \{|0\rangle, |1\rangle\}$. For instance, for a 2-qubit quantum register the computational basis is represented by $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$. We can also write the basis set for a general n-qubit register in a more compact form as $\{|x\rangle : x \in \mathbb{F}_2^n\}$ where, $\mathbb{F}_2^n$ is the set of n-bit binary numbers or strings.

Quantum Gates

We stated in the postulates, that operations or interactions with the quantum state which preserve the history (or in other words, are reversible) are the unitary transformations. *Quantum gates* are the unitary operators which are acted on the quantum

register (possibly a 1-qubit register, which is just a single qubit) to evolve the quantum state of the register and perform meaningful computation. Most of the research into design of quantum algorithms and computing is being done, to study the power of these gates which allows us to implement all the weirdness of quantum mechanics (superposition, entanglement), from the perspective of complexity (i.e how efficient the quantum computation is with respect to classical computation). This is similar to the gates we encounter in the digital electronics of a classical computer such as AND, OR, NOT etc, but in quantum setting, the gates are much more rich. Additionally, since they are unitary transformations, all computations can be 'undone' as these transformations are reversible (though some extra qubits may be required for it which do not involve in the computation explicitly. Such 'helper' qubits are called *ancillary qubits*). Physically speaking, this means that quantum computers basically don't heat up or lose energy (like classical computers) when processing data (quantum data). All quantum gates acting on n-qubit register will take all n-qubits as input and output a quantum state for n-qubit register (in other words, input and output dimensions of the Hilbert space of the quantum register is same).

Now, we briefly discuss the single and two-qubit quantum gates. The most trivial gate is the identity gate, which does nothing to the qubit.

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Similarly, n-qubit identity gate is represented by $n \times n$ identity matrix. The Pauli group itself, represent a set of quantum gates for a single qubit. The Pauli-X matrix (σ_x) or just 'X' gate (all these terms are used interchangeably) is also called the *quantum NOT* gate since, it essentially perform the classical NOT operation on the computational basis states i.e $\{|0\rangle, |1\rangle\}$.

$$X|0\rangle = |1\rangle$$

$$X|1\rangle = |0\rangle$$

The Pauli-Z matrix (σ_z) or just 'Z' gate just adds a *phase* (phase means to multiply a quantum state by a complex number whose norm is 1) of -1 to the state $|1\rangle$. Another example of single qubit gate is the *Hadamard Gate* (H) which is the unitary transformation that changes the basis from $\{|0\rangle, |1\rangle\}$ to $\{|+\rangle, |-\rangle\}$ and vice-versa.

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad H|0\rangle = |+\rangle, \quad H|1\rangle = |-\rangle$$

Note that, the Hadamard gate is it's own inverse i.e $H = H^\dagger$ and $H^2 = 1$. In fact, all elements of the Pauli group are self inverses, hence are both observables and unitary gates. Same is true for n-qubit Pauli group (n-qubit Pauli group is just the generalization of Pauli group written as $\bigotimes_{i=1}^{n} \{I, X, Y, Z\}$, shortly denoted by $\mathbb{P}_n$). By

18

linearity, any 'real' linear combination of Pauli gates also has the same property (for instance, $H = (X + Z)/\sqrt{2}$). Finally, we define another important single qubit gate which is the T gate.

$$T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\frac{\pi}{4}} \end{pmatrix}$$

We can also show that, any single qubit gate can be decomposed into the X, Y, Z and identity gates. Any unitary matrix U can be written as (using a result stated in the postulates)

$$U = exp(-i\hat{\mathbf{n}}.\overrightarrow{\sigma})$$

where, $\hat{\mathbf{n}}$ is a unit vector (in 3-dimensional Euclidean space) and $\overrightarrow{\sigma}$ is shorthand for an operator $X\hat{i} + Y\hat{j} + Z\hat{k}$ along the usual $\hat{i}, \hat{j}, \hat{k}$ directions.

Coming to the two-qubit gates (in general, multi-qubit), the gates are either applying two different single qubit gates simultaneously (for example, XX or XZ gates mean that apply X gate to both qubits simultaneously in the former and apply Z gate to first qubit and X gate on second qubit simultaneously in the latter) or are the 'controlled' versions of the single qubit gates. Essentially, these 'controlled' gates perform a 'if-then' operation. For example, CNOT (CX) (Controlled-NOT) is the 'controlled' version of NOT gate. This gate takes two qubits as input, out which one of them is the 'control' qubit and the other is the 'target' qubit. If the 'control' qubit is $|0\rangle$, then the gate does nothing to both of the qubits and if 'control' qubit is $|1\rangle$ then, it does nothing to 'control' qubit but flips the 'target' qubit.

$$CNOT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

All n-qubit quantum gates, i.e quantum gates which take n qubits as input and output form a mathematical *group* structure, formally written as $\mathbb{U}(2^n)$ called the *Unitary group* and here 'n' represents the number of qubits the operators in the group act on. It has been proved that, *universal quantum computation* (i.e any valid unitary evolution) can be realized by a set of gates called the *universal gate set* similar to the property of NAND or NOR gates being universal for classical computation. One such universal set is the *Clifford gates* + T gate. *Clifford gates* are the elements of a group called the n-qubit *Clifford group* ($\mathcal{C}_n$) which is defined to be the normalizer of the Pauli group or that the Pauli group is the normal subgroup of the Clifford group.

$$\mathcal{C}_n := \{U \in \mathbb{U}(2^n) : UPU^{\dagger} \in \mathbb{P}_n, P \in \mathbb{P}_n\}$$

. Clifford gates have been known to be efficiently implemented using the Pauli gates [?]. Also, later in this thesis, we will be working with this group extensively.

Measurement

After the application of quantum gates on the quantum register, the act of measuring the state of the register with respect to some observable is explicitly made clear by using the term *measurement gates*. Measurement, as already mentioned in the postulates, leads to 'collapse of the wavefunction' and we get as output a scalar (real number). Because of the 'irreversible' nature of measurement, usually (except in some quantum communication protocols) measurement is the last step of the quantum computation. Also, theoretically it is possible to measure only a subset of qubits in the register at the time of measurement, however in the commercial quantum computers today, the entire quantum register is measured by default (such as in IBM-Q devices). Note that, without loss of generality, we can always assume that necessary basis change has been made by applying unitary gates, so that we always measure a n-qubit register in the computational basis $\bigotimes_{i=1}^{n} \{|0\rangle, |1\rangle\}$.

There can be two utilities of the measurement, *firstly* to sample from the distribution generated as the output, over the basis states (or a subset of them) and *secondly*, to calculate a scalar expected value of a given observable in the output state of the quantum register. Often, we have to repeat a quantum computation many times (in quantum computing terminology, sometimes these repetitions are called *shots*) to generate statistics (distribution) of the basis states. Often for clarity we say that, the quantum data (unobserved quantum wavefunction or state) is collapsed to generate classical data (collapsed eigenstate's eigenvalue or some encoding of the eigenstate itself) at the time of measurement and this classical data is stored in the *classical register*. Note that, we often are interested in knowing the eigenstate into which the state of the register (quantum register) collapsed to, instead of the measurement value (eigenvalue) of the collapsed state with respect to the observable used in the measurement (Often we measure Z (σ_z) because we want to sample the distribution of the computational basis).

Quantum Circuit

The collection of the quantum register (possibly along with classical register), the unitary gates and followed by the measurement gates (of the whole register or a subset of them) is termed as a *quantum circuit*. Usually, in actual quantum circuit implementation, the register starts in the 'all up state' or 'all zero state' i.e the state $|0\rangle^{\otimes n}$ where, all the qubits in the register, are in state $|0\rangle$. The entire quantum computation is depicted in the form of a quantum circuit which is then boiled down to low level code, for the quantum computer or processor to execute *instructions* specified by the circuit. *Instruction* is similar to what we intuitively think of, given our experience with classical computers, which is any register allocation, applying unitary gates or measurement gates.

In this section, we basically reviewed the postulates of quantum mechanics and density matrix formalism of quantum systems. Next, we reviewed the elements we encounter in a typical quantum computation. In the next section, we describe about the general quantum operations, provide formal definitions and use this as a tool to study noise and decoherence in quantum systems. Next, we introduce the concept of randomized benchmarking and provide a literature review, which is fundamental for our work in this thesis.

1.2   Noise in Quantum Computation

In the previous section, we discussed mainly about two types of interactions with a quantum system, namely unitary transformations and measurement. One can say that, these interactions are determined by the user, in the sense that, we have full control over these interactions such as when do we need to apply which particular gate, when to include measurement gates or when do we have to 'reset' (transforming n-qubit register to $|0\rangle^{\otimes n}$) the register etc.

Classical computers, which are quite fault tolerant, with the probability of error (or noise) in a single bit which was 1 previously and accidentally reading it to be 0 (because of discharging of the transistor) is in the order of 10^{-9} [24]. So, in a sense we can say that, the state of a classical system (a computer or even a large enough mechanical machine) can only be changed by a user and is robust to changes in it's environment which are not in power of (or authorized by) the user. However, quantum world is different.

1.2.1   Coherent Noise

There are many sources of error (noise) when performing quantum computation. Fault tolerant computation in case of quantum computers are currently a big deal in the NISQ era. One source of error in quantum computers, is an engineering issue of calibration of parts of the computer (lasers, electronic circuits, sensors etc) that perform actual unitary gates on the quantum register. This type of noise is some unitary transformation on the register. Suppose, we want to apply multiple T gates on a qubit, but somehow, every time we give an instruction to apply T gate, the processor applies Z gate. Note that, we won't be able to get the state that we desire and there will be an error in the computation, but still register's state remains coherent (or in other words, it evolves into some other pure state) since we are always performing a unitary transformation (i.e the incorrect gate Z), although not calibrated properly. This kind of error or noise is known as *coherent error* or *coherent noise* or *unitary noise*.

1.2.2 Incoherent Noise

Quantum systems are heavily subjected to another type of interaction called *decoherence* [25, 22] due to the surrounding environment, which is another of the source of error induced in the computation. It is a non-unitary interaction with the environment, resulting from the entanglement of the quantum system with it's surrounding. The system along with the surrounding together evolve unitarily, but the reduced density matrix of the system alone turns out to be a mixed state or in other words, the density matrix of the combined system and it's surroundings taken together turns out to be non-separable. Measurement itself is a process of decoherence, however, unlike the decoherence due to the environment, it is controlled by the user and is desirable since we may want to the results of the experiment at the end of running the circuit. We have also stated in the previous section that, measurement is the last operation we perform on the register, since the interaction is irreversible. This is exactly the problem with decoherence due to environment as it can take place at any point of time (or for some duration) during computation. Often, we see that all the 'magic' in quantum algorithms happens due the coherence of the wavefunction. When a quantum system transforms into a mixed state, we are essentially losing information contained in the coherent wavefunction, to the surrounding, which hinders any potential quantum advantage. That's why, the quantum computations require very low temperatures and vacuum surrounding it so that the decoherence is as low as possible. This type of noise is also known as *incoherent noise* or *incoherent error* because of the destruction of the coherence in the wavefunction due to entanglement with the surrounding quantum state.

In this thesis, we are concerned only with the incoherent noise and it's characterization. Noise in quantum systems are studied under the framework of open quantum systems and how they evolve by possible interaction with the environment. Actual noise characterization may be hard to perform, hence only partial characterization is possible, one of the techniques for which we discuss later in this thesis. We now briefly explain the mathematical formalism to study noise (or error) induced in a quantum system, which is in interaction with the environment, thus defining *quantum channels (noise channels)* and *general quantum operations*.

1.2.3 Interactions in open quantum systems

An open quantum system is a system which is interacting with the environment i.e, is entangled with it's surrounding. The combined system and surroundings form an isolated quantum system and it evolves under unitary evolution (as stated by the postulates). But, if we consider the system alone by 'ignoring' the interaction with the environment (the interaction may not be known in full detail) then we find that the evolution of the system can be non-unitary.

Definition 9 (Quantum channels) : *Let ρ be the state of the system before inter-acting with the environment. Define a superoperator (a map that takes an operator to an operator) $\mathcal{E}$, which maps space of density operators to itself as follows:*

$$\mathcal{E}(\rho) = \sum_a M_a \rho M_a^\dagger$$

where, M_a are operators that satisfy the completeness criterion as follows:

$$\sum_a M_a M_a^\dagger = \mathbb{I}$$

where, $\mathbb{I}$ has same dimensions as ρ.

The operators M_a are known as *Kraus operators* and they 'encode' the non-unitary interaction of the system with the environment. These operators are obtained by taking the reduce trace with respect to the environment of the combined state vector for the system and environment. This type of representation of the evolution of density operators is known as the *operator-sum representation*.

Quantum channel is the form of interaction in which we are completely 'oblivious' to the outcome of interaction of the system with the surroundings, in other words, we have no information about how much the coherent information has been 'leaked' into the surroundings. This is highlighted by the fact that, the Kraus operators satisfy the completeness criterion which can be interpreted as, we are completely 'ignoring' the exact effect of the surrounding on the system and just taking an 'expected effect' on our system. This type of map is also known as *Completely Positive Trace Preserving (CPTP)* maps, since they map density operators to density operators. Noise channels are mostly studied as a CPTP maps however, more general type of noise channels also exist, which are studied as *quantum operations*. Note that, in general, the superoperator $\mathcal{E}$ maps pure state density operators to mixed state density operators, signifying the non-unitary interaction of the system with it's surroundings, but as a special case, if there is only one Kraus operator in the operator-sum representation of the action of $\mathcal{E}$ then, together with the completeness criterion, one can easily see that M_a must be unitary and hence, unitary evolution of a system is a special case of quantum channels. In particular, coherent noise can always be viewed as this special case.

Definition 10 (Operator-sum representation of coherent noise channel) : *Let $\mathcal{E}$ be a coherent noise channel and ρ be initial state of a system, then action of $\mathcal{E}$ has a operator-sum representation*

$$\mathcal{E}(\rho) = U \rho U^\dagger$$

where, U is some (approximate) unitary operator.

where, by approximate, we mean that U can be 'approximated' to sufficient degree, by a exact unitary operator $\tilde{U}$ with respect to reasonable norm (for instance, Hilbert-Schmidt norm).

We now define general quantum operations, which are the most general form of the representation of our 'knowledge' about the interaction between a system and it's environment.

Definition 11 (Quantum operations) : *Let ρ be the state of the system before interacting with the environment. Let $\mathcal{E}$ be a superoperator whose action is defined by the Kraus operators $\{M_{a\mu}\}$ (which satisfy the completeness criterion), where, we have two indices per operator. Suppose that we know the value of index a but don't know the value of index μ, so that, we only know 'partial outcome' of the interaction. Then, the effective state of the system according to our updated knowledge is given by (upto a normalization factor):*

$$\mathcal{E}_a(\rho) = \sum_\mu M_{a\mu} \rho M_{a\mu}^\dagger$$

where, $M_{a\mu}$ are operators that satisfy the inequality :

$$\sum_\mu M_{a\mu} M_{a\mu}^\dagger \leq \mathbb{I}$$

where, $\mathbb{I}$ has same dimensions as ρ.

Note that, for some Hermitian operator $A \leq \mathbb{I}$ means that, eigenvalues of A are less than or equal to 1.
We can think of the operators $M_{a\mu}$ for a fixed a, as a sort of representing the conditional effect on the system that we observe, given that we know some part of the interaction, in this case, the index a. In general, these operations do not map density operators to density operators, as they can be in general, trace decreasing because of the inequality constraint. Hence, we apply a normalization (dividing) by $Tr(\mathcal{E}_a(\rho))$ to get the final reduced density matrix of the system. If there is only one value for index a, then we essentially don't know anything about the system and we are again tracing over all possible outcomes of the interaction with the environment, thus reducing to a quantum channel. On the other hand, if there is only one value for index μ, then we essentially know the complete outcome of the interaction and hence it corresponds to measurement of the system. Therefore, this is the most general definition for interaction with a quantum system.

1.2.4 Sources of noise in quantum computation

In the previous section, we described the basic elements of a quantum computation such as the quantum circuit, gates, measurement etc. More formally, we define three different stages of computation, namely, *State preparation*, *Gate application* or the *Evolution of quantum register* and the *Measurement*.

State preparation

This stage comprises of preparing the quantum register in some pre-defined state which often, is just the default *vacuum state* or $|0\rangle^{\otimes n}$ for a n-qubit register in the computational basis. Sometimes, we may want to work or perform our computation in some different basis (for example, in the $\{|+\rangle, |-\rangle\}$ basis) or start with some specific initial state (for instance, many quantum algorithms require the register to be initially prepared in a state which is in equal superposition of the computational basis states, which is written compactly as $\frac{1}{\sqrt{2^n}} \sum_{x=0}^{2^n-1} |x\rangle$ for n-qubit register). Such transformations on the quantum register are performed in the state preparation stage. All ancillary qubits are also 'replenished' in this stage. In variational circuits, which are becoming recently popular models for quantum machine learning [11, 26], often need to encode the classical data into quantum data, which is performed in this stage.

Gate Application

In this stage, the implementation of some algorithm or model is performed, by applying gates to transform the quantum register state to some pre-defined or expected target state. The bulk of quantum computation takes place in this stage. The computation is divided into several *time steps* such that in a single time step, any number of non overlapping gates can be simultaneously performed. Suppose, CNOT gate is being applied on first and second qubits in the register in some time step, then no other gate can be performed on these two qubits in the same time step and should be postponed for the next time step. The total time steps required to perform the required computation (till possibly, measurement stage) is known as the *depth* of the quantum circuit.

There are numerous practical considerations when implementing quantum algorithms in this stage. In NISQ era, the actual quantum processors have restricted connectivity which means that, we can only apply two qubit gates to a pair of qubits which are connected according to the connectivity in the processor, therefore, we need to construct instructions for the processor accordingly. Additionally, we also need to consider the quality of qubits (on the basis of their deocoherence times) present in the processor when choosing our 'working' and 'ancillary' qubits to ensure less error in our computation. Therefore, it is crucial to perform the gate operations in an efficient manner by taking into consideration the 'quantum hardware' that the platform has to offer, in terms of connectivity and decoherence times [22, 27] of the qubits that we want to work with, for our desired application.

Measurement stage

In this stage, the quantum register is measured in preferred basis and classical data is extracted from the quantum circuit. This stage includes any unitary transformations

required to change the basis to the preferred measurement basis. Additionally, any classical processing of the data is also included in this stage.

Noise can be induced into the system at any of these stages. The errors introduced in the state preparation and measurement stages are known as *State Preparation and Measurement (SPAM)* errors. We always want to treat, the noise induced in the state preparation and measurement stages separately from the noise induced in the gate application stage. Here, we are assuming that for all practical purposes the platform/archetype is capable of 'resetting' the register to vacuum state and the measurement apparatus is sensitive enough to reduce the uncertainty in a single outcome. Hence, any model on the characterization of noise that we may wish to propose must be robust to SPAM errors, so that the error rate in actual quantum computation performed on the device by using the gates allowed by the platform/archetype (also known as *native gates*) can be studied. In this thesis, we ask, how the noise affects the computation when the depth of circuits are large. This is typical in many quantum algorithms such as Shor's algorithm, where we require many gates and 'deep' circuits to perform our required computation and we want to know the affect of noise induced as a result of 'long' computation in the gate application phase, assuming that SPAM stages are noiseless or the overall affect of SPAM errors on our computation is not of interest for our computation.

We conclude the discussion of noise in quantum computation. Ideally we may sometimes want to fully characterize the effect of noise on the quantum register which can be done through expensive techniques of quantum process tomography, since it is a statistical experiment and requires a number of trials to be done. Also, it has been shown that quantum process tomography is not robust to SPAM. Instead for NISQ devices, there has been research into partial characterization of noise by using certain figures of merit as metrics. These metrics are used to *benchmark* the 'quality' of quantum computation on a device, irrespective of platform/archetype. Such techniques have been named as 'Benchmarking Protocols' using randomization as a tool.

In the next section, we review the 'benchmarking' protocols proposed in the literature. Most of the protocols proposed, do not try to characterize gate noise in terms of it's coherence. Building onto this motivation, we propose the URB protocol, it's significance and practical implementation later in this thesis.

1.3 Randomized Benchmarking: Literature Review

In the previous section, we introduced the concept of open quantum systems and how it acts a theoretical framework and as a tool to study noise in real quantum computers. We also mentioned that, we are interested in characterizing the noise induced when applying quantum gates, therefore our theory (model) to characterize must be

robust to SPAM errors.

1.3.1   Figure of merit to characterize noise

In this thesis, we are only interested in errors induced in the computation involving arbitrarily long sequences of 'noisy' unitary gates, however, significant research has also been made to 'benchmark' or in other words, evaluating a metric that captures the performance of the quantum computing platform in a more holistic manner, that includes initial qubit state preparation errors and final measurement errors. There are mainly two aspects regarding a metric or some figure of merit, that needs to be ascertained:

firstly, that the metric that is to be evaluated must not depend on the intrinsic parameters involved in actual gate implementation in any device archetype since, we are interested in contrasting the performance of gate implementation across various platforms such as superconducting or ion-trap.

Secondly, the metric must be a good indicator of the error induced by noisy gates, for arbitrarily long computation or in other words, it must faithfully provide a value representing the error rate (though may not actually be the probability of error) irrespective of the depth of quantum circuits.

Several metrics have been proposed in recent times which are mostly clubbed under the name *Randomized Benchmarking* while, other metrics different from it have also been proposed such as Quantum Volume [28]. Today, many quantum programming software frameworks such as IBM Qiskit [29] and Forest SDK [30] provide libraries to benchmark single and two-qubit errors in their respective processor architectures which are based on randomized benchmarking (here on, referred to as the *'Standard Randomized Benchmarking' (SRB)*) though many variants have been proposed, which are claimed to be suited for different scenarios, as discussed below.

Apart from comparing and contrasting the performance of various quantum devices for computation, benchmarking of quantum devices is also vital to show empirically, that infact, fault tolerant computation is possible in NISQ devices by using the benchmark metric to show that error correction indeed improves the quality of quantum computation. Benchmarking of the performance of quantum devices also gives a perspective on the characterization and power or influence of noise on the computation, so that the errors can be better analysed and hence contribute to research in manufacturing better quantum systems with efficient architectures that allow fault tolerant computation, as is discussed below.

1.3.2 Standard Randomized Benchmarking (SRB)

Randomization has been studied as an effective tool in empirically, studying and characterising the noise associated with quantum transformations. Emerson *et al.* [31] studied circuits which apply a random sequence of noisy unitary gates followed by a reversal gate (ideally reproducing identity quantum channel) and showed that the rate of decrease of average sequence fidelity over the number of gates applied, is a strong indicator of the strength of the noise induced by the gates (assuming, that noise induced by individual gates is independent of the specific gate applied). The noteworthy aspect about this protocol of applying sequence of (Haar-random) gates is that the average noise map (average with respect to Haar measure) generated, effectively, behaves like a depolarising channel whose strength parameter λ controls the decay rate of the noise and the average fidelity of the noise channel itself is a function of this decay rate. It has been shown in [31] that average sequence fidelity for sequences of varying depth depend exponentially on the decay parameter of the depolarising channel (which essentially is the Haar-averaged noise channel). Hence, decay rate is experimentally derived (as a fit parameter) by considering gate sequences of varying depths and calculating the average sequence fidelity for each depth (by performing the aforementioned procedure of producing identity channel by applying random unitary gates followed by undoing them) and fitting an exponential curve to the plot of average sequence fidelities vs depths. The work of Emerson *et al.* [31] show that average fidelity is a good metric that characterizes the nature of the noise map. Based on this technique, Knill *et al.* [32] proposed SRB protocol and implemented it on trapped ion qubits. They used the single qubit Clifford group as the gate set from which random unitaries were chosen, motivated by the fact that Clifford unitaries can be constructed easily from Pauli gates, which constitute the native gate set (by native, we mean, specific dedicated instructions are present to implement these gates on a given quantum computer archetype and all other unitary gates have to ultimately complied into compositions of these native gates). Dankert *et al.* [33] validated their choice of gate set by showing that Clifford group is an instance of a unitary 2-design and sampling gates from any such unitary 2-design uniformly at random effectively approximates the mathematical operation of twirling which maps a quantum channel

$$\Lambda \mapsto \int dU \hat{U} \Lambda \hat{U}^{\dagger},$$

where, the integral is with respect to uniform Haar-measure over the unitary group $\mathbb{U}(2^n)$ and $\hat{U}$ is the unitary channel with respect to unitary gate U and twirling of the gate independent noise map over the unitary group exactly produces a depolarising channel as noted firstly by Emerson *et al.* in [31].

1.3.3 Interleaved Randomized Benchmarking (IRB)

SRB protocol calculates the average fidelity (infidelity) of noise channel for a gate set (preferably, a unitary 2-design) which is assumed to be invariant (or approximately invariant) for all gates in gate set. Magesan *et al.* [34] proposed *Interleaved Randomized Benchmarking* (IRB) protocol that benchmarks the error rate for a given computational gate ($\mathcal{G}$) (in contrast to a preferred gate set), again in terms of average gate fidelity similar to SRB. They propose a protocol that *firstly*, performs the SRB and extracts the decay rate (say, λ_1) and *secondly*, again performs SRB, but this time by applying the gate $\mathcal{G}$ after applying each Clifford unitary (authors mention this as 'interleaving' of the computational gate) and extracting new decay rate (say, λ_2). The required average gate fidelity of $\mathcal{G}$ is then a function of λ_1 and λ_2.

1.3.4 Real Randomized Benchmarking (RRB)

Another variant of SRB proposed by Hashagen *et al.* [35] known as *Real Randomized Benchmarking* (RRB) measures the average gate fidelity for a restricted gate set which is not an unitary 2-design. The motivation for the protocol as discussed by authors in [35] is *firstly* that, it provides more fine grained estimate for average gate fidelity when performing computations that result only in real valued density operators (in other words, when one is interested only in unitary transformations which do not induce complex parts to the values in the input density operators) and *secondly*, when working with an error correction scheme, the whole Clifford group may not be accessible under the scheme and hence, to benchmark the performance of such a scheme one has to work with a restricted gate set, as noted by the authors, for example, for [4,2,2] code in [36].

Brown *et al.* [37] also claim that, it is certainly sufficient but not necessary for the gate set to be a unitary 2-design to conclusively benchmark error rates through similar procedures as prescribed in SRB, rather subgroups of a unitary 2-design like the real Clifford group (used in RRB as proposed in [35]) which is a subgroup of the Clifford group can provide approximate average gate fidelities within a factor of two or less. Aforementioned protocols benchmark the performance of computation with respect to a chosen gate set and independent of the actual implementation of gates in a practical device, hence, the protocols do not take into consideration the errors specifically introduced due to compilation of gates in the chosen gates into the native gates allowed by the processor and evidently, may underestimate the performance of an actual processor due to sub optimal compilation of circuits.

1.3.5 Direct Randomized Benchmarking (DRB)

We finally mention *Direct Randomized Benchmarking* (DRB) proposed by Proctor *et al.* [38] to benchmark the performance of gate set made of native gates and not of a unitary 2-design (or it's subgroup) and hence, directly providing the error rate of

gates which are directly implemented in the quantum computing processor (and hence is archetype dependent). Still, we note that the original SRB protocol is prevalent in the error correction literature and as a practical standard for benchmarking error rates of gates.

In this chapter, we introduced the basic mathematical tools and provided intuition to understand the problem at hand, mainly to characterize noise in actual quantum computers which can be treated as a figure of merit to compare between platforms and archetypes. We specifically looked at some prevalent techniques for partial characterization using benchmarking protocols.

In the next chapter, we discuss at length about a variant of SRB known as the *Unitarity Randomized Benchmarking* (URB) which highlights the type of error (coherent or incoherent error) introduced by the noise.

Chapter 2

Unitarity Randomized Benchmarking (URB)

2.1   Motivation

In SRB and other variants discussed above, the primary metric of interest was the average gate infidelity of a gate set (may or may not be native gate set) or of a single computational gate, which was interpreted as a strong indicator for the probability of error per gate chosen randomly from our gate set. However, we can further characterize the noise, distinct from quantifying the error introduced by the map, in terms of the coherence of the error. For measuring the actual error rates, as discussed in threshold theorems for fault tolerant quantum computation, the metric of interest to quantify the error rate is the diamond distance between the ideal quantum channel and actual channel [39]. It has been shown, that the upper bound on the worst case error scales as square root of the average gate infidelity and is saturated by unitary noise channel and in the case of Pauli channels it scales linearly [40]. Since, the actual error per gate scales with average fidelity differently for different channels, based on the nature of the channel, it is imperative to characterize the type of channel in order to claim the effectiveness of the average infidelity to quantify the gate error precisely. If the noise channel is (approximately) a unitary channel, then the errors induced in the density operators are termed to be (approximately) coherent, for example, for single qubit gates applied to single qubit pure state, unitary noise correspond to over or under rotation of the state vector on the surface of Bloch sphere. Note, that unitary noise channels map pure states to pure states since they are both trace preserving and unital maps and hence, the state of the qubit(s) remain in a 'coherent' state. On the other hand, if the noise channel is not a unitary channel (or rather, is far from being an ideal unitary channel; such a map may be non-unital or non-trace preserving or both) for instance, a depolarising channel then, the errors introduced by the noisy gates are defined to be 'incoherent'. In the case of a single qubit, a depolarising channel maps pure states to mixed states, thereby reducing the norm of the state

vector, in the Bloch sphere representation. Wallman *et al.* [41] define unitarity as a metric that captures intuitively, the idea of the extent to which the noise associated with a unitary gate is itself not unitary. They propose a protocol similar to SRB that benchmarks this metric for a noise channel attached to noisy gates, with respect to a gate set which is a unitary 2-design.

In this thesis, we've tried to implement a slightly modified version of the protocol mentioned in [41], which is more suited for experimental purposes, as proposed by Dirkse *et al.* [42]. We first introduce key concepts related to URB followed by a discussion on what URB actually measures and it's significance in complementary with SRB and possibly, it's aforementioned other variants. Our main experimental results are summarised in next chapter, where we first present our results on simulations of URB protocol for various noise channels using Aer simulator in IBM-Qiskit and it's consistency with the theoretical results. Finally, we present the unitarity benchmark results for simulated noise maps for native gates present in two of the IBM-Q devices, *burlington* (display_name: ibmq_burlington) [43] and *melbourne* (display_name: ibmq_16_melbourne) [44], which are 5-qubit and 15-qubit processors respectively.

2.2 Concepts

We begin this section by introducing (revisiting) important concepts to understand URB protocol namely, purity and unitarity. While purity is essentially the Hilbert-Schmidt norm of a density operator, unitarity of a noise channel is the figure of merit, estimated by the URB protocol which we define below and highlight it's similarity with the concept of purity.

2.2.1 Purity

Purity of a quantum state (of qubits) which when expressed as a density operator, ρ (a self adjoint operator with $\mathbf{Tr}[\rho] = 1$) is given by, $\mathrm{Tr}(\rho^\dagger \rho) \in [0, 1]$. Note, that $\mathbf{Tr}[\rho^\dagger \rho] = 1$ if and only if ρ is a pure state or in other words, has a state vector representation in the Hilbert space $\mathcal{H}$ of dimension $d = 2^n$ where, n is the number of qubits.

2.2.2 Unitarity

Let $\mathcal{E}$ be a noise channel which is a superoperator that maps the space of quantum states (which are represented by density operators over Hilbert space $\mathcal{H}$) to itself. Note, that for the sake completeness, here we assume that noise channels are non-trace-increasing and in particular, a noise channel can be trace decreasing so, we define $\mathcal{E}$ as a general quantum operation. A natural candidate to quantify the coherence of a noise channel (loosely speaking, the extent to which it behaves as a unitary

transformation) or unitarity can be $\int d\psi \mathbf{Tr}[\mathcal{E}(\psi)^\dagger \mathcal{E}(\psi)]$ where, the integral is with respect to uniform Haar measure over the space of pure states in $\mathcal{H}$. In other words, it is the average purity of the output state, averaged over all possible input pure states. But, as Wallman *et al.* [41] argue that this definition of unitarity may not work for general quantum operations and not just CPTP (Completely Positive Trace Preserving) maps. The general definition for unitarity proposed by Wallman *et al.* in [41, Eq. 4] is

$$u(\mathcal{E}) = \frac{d}{d-1} \int d\psi \mathbf{Tr}[\mathcal{E}'(\psi)^\dagger \mathcal{E}'(\psi)]. \tag{2.1}$$

Here, $\mathcal{E}'(\psi) = \mathcal{E}(\psi) - \frac{\mathbf{Tr}[\mathcal{E}(\psi)]}{\sqrt{d}}\mathbb{I}$. However, for this paper we define unitarity as in [42] (Note, in this definition $\mathcal{E}$ is a CPTP map and hence, we differentiate it from a quantum operation by calling it a quantum channel):

$$u(\mathcal{E}) = \frac{d}{d-1} \int d\psi \mathbf{Tr}\left\{ \left[\mathcal{E}\left(|\psi\rangle\langle\psi| - \frac{\mathbb{I}}{d} \right) \right]^2 \right\} \tag{2.2}$$

where, the integral is again with respect to uniform Haar measure over the space of pure states in $\mathcal{H}$. It is noteworthy, that in order to convert unitarity as a normalized metric (i.e, take values in $[0,1]$, similar to average gate fidelity in SRB and other variants) the constant factor $\frac{d}{d-1}$ has been included. An equivalent definition of unitarity as noted in [15,16] is:

$$u(\mathcal{E}) = \frac{1}{d^2-1} \sum_{\sigma,\tau \in \mathbb{P}^*} \left\{ \mathbf{Tr}[\tau\mathcal{E}(\sigma)] \right\}^2 \tag{2.3}$$

Here, $\mathbb{P}^*$ is the normalized Pauli group, excluding the normalized identity (the normalization is with respect to Hilbert-Schmidt norm, $||\sigma||_2 = \sqrt{\mathbf{Tr}[\sigma^\dagger \sigma]}$). This is the form of unitarity that we use in it's estimation protocol through URB as suggested in [42]. We discuss the interpretation and possible merits and demerits of this form later.

2.3 The protocol

Now, we formally state the URB protocol followed by a discussion on the user defined parameters used in it. Authors in [41, 42] propose two equivalent implementations of URB known as the *single copy* and *two copy implementation*. We remark here, that we follow *single copy implementation* in our experiments and hence, we describe here only the *single copy implementation*. We suggest the reader to refer [42] for the *two copy implementation* and it's equivalence and comparison with *single copy implementation*.

> **Input**: Gate set $\mathbb{G}$, sequence lengths (depths) $\mathbb{M}$, Number of iterations (or number of random sequences) $\mathbf{N}_m$ for every depth $m \in \mathbb{M}$.
> **Begin**:
>   **For all** depth $m \in \mathbb{M}$:
>   **Repeat $\mathbf{N}_m$** times:
>   Sample a sequence of m gates $\mathcal{G}_{j_1}, \mathcal{G}_{j_2}, ..., \mathcal{G}_{j_m}$ uniformly at random from $\mathbb{G}$.
>   Compose the sequence $\mathcal{G}_{\mathbf{j}} = \mathcal{G}_{j_m}...\mathcal{G}_{j_2}\mathcal{G}_{j_1}$.
>   **For all** non-identity Pauli's P, Q:
>   Prepare $\rho^{(P)} = \frac{I+P}{d}$ and $\hat{\rho}^{(P)} = \frac{I-P}{d}$. Apply $\mathcal{G}_{\mathbf{j}}$ to each state. Measure $E^{(Q)} = Q$ large number of times. Estimate shifted purities, $q_{\mathbf{j}} = \frac{1}{d^2-1} \sum_{P,Q \neq I} \{\mathbf{Tr}[E^{(Q)}\mathcal{G}_{\mathbf{j}}\rho^{(P)}] - \mathbf{Tr}[E^{(Q)}\mathcal{G}_{\mathbf{j}}\hat{\rho}^{(P)}]\}^2$
>   Average over all sequences $\tilde{q}_m = \frac{1}{\mathbf{N}_m} \sum_{\mathbf{j}} q_{\mathbf{j}}$
>   Fit exponential curve $\tilde{q}_m = Bu^{m-1}$ where, u is unitarity and B is a constant absorbing SPAM.
> **End**

Algorithm 2.3. Pseudocode for URB Protocol (single copy implementation)

We remark here, that the actual implementation of the protocol is slightly different from the one mentioned above, a discussion of the precise experimental implementation is included in subsequent section. Below, we discuss the parameters used in the protocol mentioned in Algorithm 2.3 above.

Gate set ($\mathbb{G}$)

The set of gates from which random gates are sampled and applied to the initial state. The motivation for defining a gate set for choosing random gates, is similar in spirit, to that of averaging the noise channel over all possible input states, as described in SRB protocol. Wallman *et al.* in [41] mention that the gate set to be preferably a unitary 2-design, so that the integral as mentioned in Eq. (2.1) can be effectively approximated through experiments. We remark here, that in [41, 42] as well as in this paper, $\mathcal{E}$, or the noise channel in question, is gate-independent. For our results, we choose our gate set to be the Clifford group (single and two-qubit Clifford groups for estimating unitarity of noise maps of single and two-qubit gates respectively).

Sequence lengths (depths) ($\mathbb{M}$)

An array of number of unitary operators from the $\mathbb{G}$ that has to be chosen to empirically estimate the unitarity. Ideally, it is recommended to use many values for m so that the error introduced in the estimate of unitarity, (due to variance in the

experimental results, between two different depth values) when fitting the curve, can be minimized.

Number of iterations ($\mathbf{N}_m$)

For each depth chosen in the $\mathbb{M}$, in each iteration, the average sequence purities are calculated for $\mathbf{N}_m$ different sequences sampled from $\mathbb{G}$ and they are averaged to get the empirical average purity $\tilde{q}_m$ for given depth m. Again, it is recommended to choose high enough $\mathbf{N}_m$ so that between sequence variance $\mathbb{V}[q_{\mathbf{j}}]$ is minimized. Dirkse *et al.* in [42] gave a theoretical lower bound on $\mathbf{N}_m$ for given confidence parameters ϵ and δ, where, ϵ can be treated as the accuracy or the deviation of the empirical average $\tilde{q}_m$ from $\mathbb{E}[q_{\mathbf{j}}]$ and δ as upper bound on the probability that $\tilde{q}_m$ is 'more than ϵ away' from $\mathbb{E}[q_{\mathbf{j}}]$, which is mathematically expressed by the concentration inequality,

$$\Pr[|q_{\mathbf{j}} - \mathbb{E}[q_{\mathbf{j}}]| \geq \epsilon] \leq \delta.$$

This concludes a high level discussion on the URB protocol. Here, we would like to mention two comments, *firstly*, that Wallman *et al.* in [41] derive that the unitarity, similar to average gate fidelity, decays exponentially with increasing number of gates in the gate sequence applied given by, $\mathbb{E}[q_{\mathbf{j}}] = A + Bu^{m-1}$. Dirkse *et al.* in [42] propose a modification, to prepare two separate initial states ρ and $\hat{\rho}$ and apply the same sequence $\mathcal{G}_{\mathbf{j}}$ to each of the states simultaneously, so that we effectively work with a traceless density operator $\tilde{\rho} := \frac{1}{2}(\rho - \hat{\rho})$ that changes the form of the fit curve to $\tilde{q}_m = Bu^{m-1}$, which is easier to fit since, by taking the log of the empirical average purities and plotting them against depths we can reduce it to a linear regression problem.

Secondly, the mathematical form of the two initial states ($\rho^{(P)}$ and $\hat{\rho}^{(P)}$) to be prepared, as mentioned in the single copy URB implementation mentioned previously, is advocated in [42]. The authors argue that the SPAM absorbing constant B in the fit model $\tilde{q}_m = Bu^{m-1}$ has the property that $|B| \leq 1$ [42, Lemma 11.] and since, we would ideally want to maximize the 'signal strength' i.e, the experimental values of average sequence purities in case, for the channels which are vastly incoherent ($0 \leq u \leq 0.5$), we want to choose our initial states and measurement operators so that the $B = 1$ and the choice of $\rho^{(P)} = \frac{I+P}{d}, \hat{\rho}^{(P)} = \frac{I-P}{d}$ and $E^{(Q)} = Q$ (for single copy implementation) is one such choice for which $B = 1$. Another reason for this choice, is the fact that, Pauli operators are natural observables, for which the we can easily calculate the expectation values (i.e, the trace terms inside the expression for shifted purities) and though the initial states proposed are mixed states for $n \geq 2$ qubits, the mixed state can be easily expressed as convex sum of pure states that can be easily prepared in actual implementation, as is discussed in detail later.

2.4 Significance of URB

In this section, we discuss how exactly the definition of unitarity in Eq.s 2.1, 2.2, 2.3, captures the deviation of a noise channel from being a unitary channel. Specifically, we will focus on the form of unitarity defined in Eq. 2.3 and establish the theoretical correctness of URB protocol, specifically by drawing similarity between Eq. 2.3 and the definition of shifted purities in the aforementioned protocol. Next, we compare URB with SRB and it's other variants and conclude by highlighting the role of URB, complementary to SRB and other variants in the theory of benchmarking and error correction.

Let, $\mathcal{E}$ be a CP (Completely Positive) map from the space of quantum states in $\mathcal{H}$ to itself. A density operator ρ can be expanded in the Pauli operator basis $\{\sigma_i\}_{i=0}^{d^2-1}$ as follows: (where, d is the dimension of $\mathcal{H}$ and for now, $d = 2^n$, n being the number of qubits)

$$\rho = \sum_{i=0}^{d^2-1} \rho_i \sigma_i \tag{2.4}$$

We can then write $\mathcal{E}$ as a $d^2 \times d^2$ matrix. Assuming, that the $\mathcal{E}$ is also trace preserving (note, it is not a restrictive assumption for our case), the matrix for $\mathcal{E}$ looks like:

$$\mathcal{E} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \vdots & & \ddots & & \\ \mathcal{E}_n & & & \mathcal{E}_u & \\ \vdots & & & & \ddots \end{pmatrix}$$

Here, $\mathcal{E}_n$ (of dimension $d^2 - 1 \times 1$) is the non-unital part of $\mathcal{E}$, since, if this vector is non-zero then the map $\mathcal{E}$ is said to be non-unital, i.e, doesn't map identity (σ_0) to itself. On the other hand, $\mathcal{E}_u$ (of dimension $d^2 - 1 \times d^2 - 1$) is the unital part of $\mathcal{E}$. It has been shown that, $\mathcal{E}_u$ is orthogonal matrix (for, real quantum channels i.e, $\mathcal{E}$ maps Hermitian operators to Hermitian operators) if and only if $\mathcal{E}$ is an unitary channel or mathematically, $\mathcal{E}$ is of the form $U\rho U^\dagger$ for some unitary operator U over the Hilbert space $\mathcal{H}$ [39]. Wallman *et al.* in [41, Proposition 1] prove that the definition of unitarity defined in Eq. 2.1 is equivalent to:

$$u(\mathcal{E}) = \frac{\mathbf{Tr}[\mathcal{E}_u^\dagger \mathcal{E}_u]}{d^2 - 1} \tag{2.5}$$

Here, we note that, if the noise channel $\mathcal{E}$ is unitary then, since, $\mathcal{E}_u$ has to be orthogonal, we see that $u(\mathcal{E}) = 1$. The converse of this statement is also true and in fact, for channels other than unitary channels, Wallman *et al.* in [41, Proposition 7] show that $u(\mathcal{E}) < 1$. We establish that, in fact, unitarity is a viable figure of merit, to measure

the deviation in the behaviour of a noise channel from being a unitary channel.

We now discuss, how the URB protocol mentioned in 2.3, lets us calculate empirically, the unitarity for arbitrary noise channels. By a slight algebraic manipulation, the equivalence of the form of Eq. 2.3 and Eq. 2.5 can be established as follows. *Firstly*, note that, with respect to form of a density operator mentioned in Eq. 2.4, we can write the action of $\mathcal{E}$ on ρ in an algebraic manner as:

$$\mathcal{E}(\rho) = \sum_{i,j}(\mathcal{E}_{ij}\rho_j)\sigma_i$$

Now, Eq. (3) can be written as:
$$u(\mathcal{E}) = \frac{1}{d^2-1}\sum_{\sigma,\tau\in\mathbb{P}^*}\left\{\mathbf{Tr}[\tau\mathcal{E}(\sigma)]\right\}^2$$

$$= \frac{1}{d^2-1}\sum_{i,j=1}^{d^2-1}\left\{\mathbf{Tr}[\sigma_j\mathcal{E}(\sigma_i)]\right\}^2$$

Using Eq. (7),

$$= \frac{1}{d^2-1}\sum_{i,j=1}^{d^2-1}\left\{\mathbf{Tr}\left[\sigma_j\sum_{k=0}^{d^2-2}\mathcal{E}_{u_{ki}}\sigma_k\right]\right\}^2$$

$$= \frac{1}{d^2-1}\sum_{i,j=1}^{d^2-1}\left\{\sum_{k=0}^{d^2-2}\mathcal{E}_{u_{ki}}\mathbf{Tr}[\sigma_j\sigma_k]\right\}^2$$

Since, $\mathbf{Tr}[\sigma_j\sigma_k] = \delta_{jk}$, (where, $\delta_{ab} = 1$ if and only if $a = b$ and is zero otherwise and using the fact that, $\sigma_i \in \mathbb{P}^*$ (normalized Pauli group) $\forall i$)

$$= \frac{1}{d^2-1}\sum_{i,j=0}^{d^2-2}(\mathcal{E}_{u_{ji}})^2$$

$$= \frac{1}{d^2-1}\mathbf{Tr}[\mathcal{E}_u^{\dagger}\mathcal{E}_u]$$

(where, the last equality is true for real noise channels).

Now, we show that the expression for shifted purities specified in the URB protocol is equal to the unitarity of composed noise map of the sequence of unitaries applied, upto a constant factor. By slightly modifying the expression for shifted purities given by:

$$q_{\mathbf{j}} = \frac{1}{d^2-1}\sum_{P,Q\neq I}\left\{\mathbf{Tr}[E^{(Q)}\mathcal{G}_{\mathbf{j}}\rho^{(P)}] - \mathbf{Tr}[E^{(Q)}\mathcal{G}_{\mathbf{j}}\hat{\rho}^{(P)}]\right\}^2$$

$$= \frac{1}{d^2-1}\sum_{P,Q\neq I}\left\{\mathbf{Tr}[Q\mathcal{G}_{\mathbf{j}}(\frac{\mathbb{I}+P}{d})] - \mathbf{Tr}[Q\mathcal{G}_{\mathbf{j}}(\frac{\mathbb{I}-P}{d})]\right\}^2$$

$$= \frac{4}{d^2-1}\sum_{P,Q\neq I}\left\{\mathbf{Tr}[Q\mathcal{G}_{\mathbf{j}}(\frac{P}{d})]\right\}^2$$

More precisely, since, P and Q here are ordinary Pauli gates, by normalizing them (with respect to Hilbert-Schmidt norm), the final expression for shifted purities looks like,

$$= \frac{4}{d^2-1} \sum_{\sigma,\tau \in \mathbb{P}^*} \left\{ \mathbf{Tr}[\tau \mathcal{G}_\mathbf{j}(\sigma)] \right\}^2$$

Here, we see that, shifted purities for sequence $\mathcal{G}_\mathbf{j}$ is equal to the unitarity of the noise map induced by applying the sequence (assuming, $\mathcal{G}_\mathbf{j}$ is noisy) upto a constant factor. We note here that, this constant factor is consumed in the constant B when extracting u by fitting the exponential curve. We also mention the intuitive reason behind the curve fitting step in the protocol, that the empirical average purity for given sequence length correspond to the unitarity of the noise channel (upto constant factor) composed 'length of the sequence many' times and thus the unitarity is raised to power of the sequence length. We therefore, conclude the significance and correctness of URB protocol as proposed in [41] with modification in state preparation and measurement for single copy implementation suggested in [42].

Now, we try to compare and contrast URB protocol with SRB protocol and it's other variants, by carefully inspecting the significance of each of it's metrics and how they are related. *Firstly*, in SRB protocol, after the sequence of gates (preferably from a unitary 2-design) are applied, an inverting gate is applied, so that ideally the quantum operation is equivalent to the identity channel, since, we were interested in the average survival probability of a given input state, under the action of the averaged noise map of the gate set. Hence, we require efficient implementation (and reversal) of unitary gates in the gate set, since, errors in compiling inefficient circuit implementations might affect the average infidelity. However, we see that, in URB protocol, there is no such need for an inverting gate and the gates can be applied on the fly, without concerning about storing the gates being applied in the memory, in order to implement the inverting gate, as is required in SRB.
Secondly, SRB protocol can be implemented theoretically, with any arbitrary pure state as the input state and calculating it's survival probability, whereas, we notice that in URB (both for single copy and two copy implementations) there exist optimal input state and measurement operator, so that the shifted purity values are of maximum strength (i.e, they are sufficiently larger than 0) [42].
Thirdly, in SRB, we experimentally determine the probability that an arbitrary input state 'survives' the noisy channel, averaged over many iterations, as compared to, the expectation values of Pauli operators for the output states (by output states, we mean, the resulting state, after applying the sequence of noisy gates) in URB. In this paper, we discuss URB with the assumption of gate-independent Markovian noise, but we suspect that it can be easily extended to gate-dependent Markovian noise, as with SRB.

Wallman *et al.* in [41, Proposition 8] deduce the following upper bound on average gate fidelity F_{avg}, in terms of unitarity of the average noise map of the gate set , u:

$$\left(\frac{dF_{\text{avg}} - 1}{d - 1}\right)^2 \leq u$$

This inequality shows that, if u is strictly less than 1, then the maximum (non-trivial) average gate fidelity for the gate set can be $\frac{\sqrt{u}(d-1)+1}{d}$, but if the noise is unitary, i.e, $u = 1$ then, there is no non-trivial upper bound for the average gate fidelity and indeed, with unitary noise, the gate fidelity may be close to zero. Therefore, we see that, gate fidelity alone is not sufficient to characterize and quantify the performance of a noisy computation, since a low value for average gate fidelity with respect to a non-unitary noise map, may be the best case benchmark under given, operating characteristics (noisy gate implementations in the quantum processor under consideration) of the computation which, directly affects the unitarity of the noise map. Hence, URB protocol must be treated as a complementary procedure, along with SRB, in order to safely benchmark the quality of quantum computation. In quantum error correction literature and specifically with respect to threshold theorems for fault tolerant computation, diamond distance of the noise map with respect to the identity channel is regarded as the proper metric for noise strength. Mathematically, the diamond norm of a superoperator $\mathcal{E}$ is defined as:

$$||\mathcal{E}||_\diamond = \max_\rho ||\mathcal{E} \otimes \mathcal{I}(\rho)||_1 \tag{2.6}$$

where, $\mathcal{E}$ acts on $\mathcal{H}$ and $\mathcal{I}$ is the identity map acting on $\mathcal{H}'$ and maximization is with respect to density operators on the combined Hilbert space $\mathcal{H} \otimes \mathcal{H}'$. The diamond distance of a noise channel $\mathcal{N}$ from the identity channel $\mathcal{I}$ is then given by [39]:

$$D_\diamond(\mathcal{N}) := ||\mathcal{N} - \mathcal{I}||_\diamond \tag{2.7}$$

We note here, that the diamond distance quantifies the worst case error rate of a noise channel with respect to the identity channel and the average gate infidelity quantifies the average error rate of the noise channel. Experimentally estimating the diamond distance is hard, but it can be bounded using experimental benchmarking protocols. Kueng *et al.* [45], gave the following bounds on the diamond distance $D_\diamond$ of a noise channel $\mathcal{N}$, whose average gate infidelity, $r(\mathcal{N}) = \frac{(d-1)(1-p)}{d}$, where, p is the benchmarking parameter (fit parameter) in SRB and unitarity of $\mathcal{N}$ is u:

$$\frac{\sqrt{(d^2 - 1)(1 - 2p + u)}}{2d} \leq D_\diamond \leq \frac{d\sqrt{(d^2 - 1)(1 - 2p + u)}}{2} \tag{2.8}$$

We would like to conclude this section by an important corollary of the above inequality, that if the noise channel is unitary, then the worst case error rate of the noise channel, $D_\diamond$ scales as $\sqrt{r(\mathcal{N})}$ whereas, if the channel is far from unitary, for

instance a depolarising channel, with depolarising parameter p, $D_\diamond$ scales as $r(\mathcal{N})$. Therefore, unitarity and infidelity together can provide insights into worst case error rate of a noise channel which is the theoretical metric used for benchmarking fault tolerant computation.

In this chapter, we introduced the concepts related to Unitarity Randomized Benchmarking (URB) and provided it's significance. We argued that, URB should be treated as a complementary procedure along with SRB in effectively bounding the worst case error rate of a noise channel. In the next chapter, we present the results of our work, followed by it's discussion.

Chapter 3

URB Simulations and Experiments in IBM-Q devices using Qiskit

3.1 Overview

We now present, our work of programming the URB protocol *(single copy implementation)* in Qiskit. *Firstly*, we mention in detail, our implementation of modified version of URB as proposed in [41, 42] in Qiskit. *Secondly*, we discuss the results we obtained via simulations of various well defined noise channels, in particular, (completely) depolarising channel and single qubit bit-flip noise channels, using Aer simulator in Qiskit. *Thirdly*, we showcase the consistency of our experimental results with that of predicted values with sub-percent implementation errors. *Finally*, we discuss our results on benchmarking the noise map of single and two-qubit gates in actual IBM-Q devices (5-qubit [43] and 15-qubit [44]) by simulating noise in these devices.

Here, we again mention for completeness, that we use Clifford group as our gate set and implemented the protocol for single and two-qubit Clifford unitaries. For implementing a Clifford gate in our protocol, we use the library QuaEc: Quantum Error Correction Analysis in Python [46], for generating the single and two-qubit Clifford group and randomly sampling a Clifford unitary, decomposed into basic gates namely, *I, X, Y, Z, H, T,* Controlled-X, Controlled-Z and SWAP gates. Using the functions available in the library, we extract QASM (Quantum Assembly Language) [47] string, specifying the decomposition and implement it manually in Qiskit.

3.2 Details of URB Implementation in Qiskit

In this subsection, we first explain our implementation of the state preparation for single and two-qubit URB implementation as prescribed in Algorithm 2.3. Next, we describe how we experimentally calculate the shifted purity, by calculating the expectation with respect to normalized Pauli operators, mentioned in Algorithm 2.3.

3.2.1 Single and two-qubit state preparation for benchmarking

As mentioned in the URB protocol in Algorithm 2.3, in each iteration of given depth, we first sample a sequence of depth many Clifford unitaries using QuaEc. After which, we prepare two states, $\rho^{(P)} = \frac{(I+P)}{d}$ and $\hat{\rho}^{(P)} = \frac{(I-P)}{d}$, for each non identity Pauli gate P over n qubits.

For single qubit case, there are only three choices for P namely, X,Y,Z gates (Pauli X, Y and Z respectively) and the states. $\rho^{(P)}$ and $\hat{\rho}^{(P)}$ are respectively the density matrices corresponding to positive and negative eigenstates of the respective Pauli operator P. For instance, if $P = X$ then, $\frac{(I+X)}{2} = |+\rangle\langle+|$ and $\frac{(I-X)}{2} = |-\rangle\langle-|$ where, $|+\rangle$ and $|-\rangle$ are respectively the positive and negative eigenstates of Pauli X. So for single qubit, for each choice of P, we initialize two separate circuits, one for each input state and apply the same randomly generated sequence of Clifford unitaries to both of the circuits. For $n > 1$, both $\rho^{(P)}$ and $\hat{\rho}^{(P)}$ are mixed state density matrices, as $\text{Tr}\left[\frac{(I+P)(I+P)}{d^2}\right] = \text{Tr}[\frac{2I}{d^2}] = \frac{2}{d} < 1$, ($P^2 = I, \text{Tr}[I] = d$ and $\text{Tr}[P] = 0$ for all non identity Pauli gates P). Similarly, $\text{Tr}\left[\frac{(I-P)(I-P)}{d^2}\right] = \text{Tr}[\frac{2I}{d^2}] = \frac{2}{d} < 1$.

For two-qubit state preparation, we decompose the mixed states into convex sum of pure state density matrices, for instance, if $P = X \otimes X$,

$$\frac{I + X \otimes X}{4} = \frac{1}{4}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix},$$

$$= \frac{|\psi^+\rangle\langle\psi^+| + |\phi^+\rangle\langle\phi^+|}{2},$$

where $|\psi^+\rangle = \frac{|00\rangle+|11\rangle}{\sqrt{2}}$ and $|\phi^+\rangle = \frac{|01\rangle+|10\rangle}{\sqrt{2}}$ and

$$\frac{I - X \otimes X}{4} = \frac{1}{4}\begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix},$$

$$= \frac{|\psi^-\rangle\langle\psi^-| + |\phi^-\rangle\langle\phi^-|}{2}.$$

where, $|\psi^-\rangle = \frac{|00\rangle-|11\rangle}{\sqrt{2}}$ and $|\phi^-\rangle = \frac{|01\rangle-|10\rangle}{\sqrt{2}}$.

Similarly, for each non identity Pauli gate P, $\rho^{(P)}$ and $\hat{\rho}^{(P)}$ each can be written as sum of two pure state density matrices, so for two-qubit case, we initialize four circuits per choice of P, where the input state for each circuit is one of each pure states that make

up the mixed states $\rho^{(P)}$ and $\hat{\rho}^{(P)}$. For example, we initialize two circuits with initial states prepared in the Bell states, $|\psi^+\rangle$ and $|\phi^+\rangle$, apply the same random sequence to each of the circuits to carry out the computation for $\rho^{(P)}$. Similarly, initialize two circuits with Bell states $|\psi^-\rangle$ and $|\phi^-\rangle$ as initial states and apply the same sequence to each of them, to simulate the step for $\hat{\rho}^{(P)}$. We note here, that the factor of 0.5 in the representation of $\rho^{(P)}$ and $\hat{\rho}^{(P)}$ in terms of pure states, is taken into account in the measurement step, as discussed below.

3.2.2 Calculating the expectation of Pauli observable

For completeness, we state here, the expression for shifted purity:

$$q_{\mathbf{j}} = \frac{1}{d^2 - 1} \sum_{\mathrm{P,Q} \neq \mathrm{I}} \left\{ \mathrm{Tr}[E^{(Q)}\mathcal{G}_{\mathbf{j}}\rho^{(P)}] - \mathrm{Tr}[E^{(Q)}\mathcal{G}_{\mathbf{j}}\hat{\rho}^{(P)}] \right\}^2.$$

Here, $E^{(Q)} = Q$, where Q is any non identity Pauli gate. We see that the term in the summation is nothing but the square of the difference in the expected values of the Pauli observable Q in the states resulting, after the application of the random sequence of gates to the input states $\rho^{(P)}$ and $\hat{\rho}^{(P)}$.

In our implementation, for all circuits initialized in the previous state preparation step in Sec 3.2.1, we calculate the expected value of Q. It is known that, since any Pauli observable has only two eigenspaces, corresponding to it's two possible values after measurement viz. +1 and -1, we then, essentially perform a positive operator valued measurement (POVM). If POVM element corresponding to positive eigenspace of Q is M, then we can write $Q = 2M - I$ and then $\mathrm{Tr}[Q\rho] = 2\mathrm{Tr}[M\rho] - 1$, where $\mathrm{Tr}[M\rho]$ is nothing but the probability that the state ρ collapses to positive eigenspace of Q. In our implementation, we always measure in the computational basis, hence, we first rotate the basis as follows. Mathematically, we can write,

$$\mathrm{Tr}[Q\rho] = \mathrm{Tr}[U^\dagger \tilde{Z} U \rho] = \mathrm{Tr}[\tilde{Z} U \rho U^\dagger],$$

where U is a unitary operator that satisfies $U^\dagger \tilde{Z} U = Q$ and the second equality corresponds to change of basis transformation to $\tilde{Z}$ (a n-qubit Pauli observable) basis. Note that, for single qubit case, $\tilde{Z} = Z$ (Pauli Z) and for two-qubit case, $\tilde{Z}$ here, represents the Pauli observables $Z \otimes I$ for cases $Q = X \otimes I$ and $Q = Y \otimes I$, $I \otimes Z$ for cases $Q = I \otimes X$ and $Q = I \otimes Y$ and $Z \otimes Z$ for cases $Q = X \otimes X$ and $Q = Y \otimes Y$. Hence, we empirically calculate the probability that the state collapses to $|0\rangle$, for single qubit implementation and for two-qubit implementation of the protocol, depending on the observable Q, we consider only the subset of computational basis states spanning the positive eigenspace of Pauli observable $\tilde{Z}$ as mentioned above and calculate the empirical probability of the state collapsing into one of these basis states.

For instance, if $Q = X$, then $U = H$ (Hadamard Gate), as it is known that $HZH = X$, so we apply an additional Hadamard gate to each of the circuits and

then count the number of instances, the quantum register in each circuit collapses in $|0\rangle$ ($|00\rangle$). Considering single qubit implementation for this example, let p_1 and p_2 be the empirical counts we obtained for circuits initialized with $\rho^{(P)}$ and $\hat{\rho}^{(p)}$ respectively, then $\text{Tr}[X\mathcal{G}_\mathbf{j}\rho^{(P)}] = 2p_1 - 1$ and $\text{Tr}[X\mathcal{G}_\mathbf{j}\hat{\rho}^{(P)}] = 2p_2 - 1$ for some choice of P and sequence $\mathcal{G}_\mathbf{j}$. As mentioned before, for single qubit case, both $\rho^{(P)}$ and $\hat{\rho}^{(p)}$ are pure states which can be directly implemented and hence we have to calculate the expectation of Q for two circuits for every choice of pair (P, Q), but for two-qubit case, we calculate the expectation of Q for the 2 circuits initialized in the two pure states appearing in the mixed state corresponding to $\rho^{(P)}$ and $\hat{\rho}^{(p)}$ and then multiply by the factor 0.5, as each of $\rho^{(P)}$ and $\hat{\rho}^{(p)}$ are equal mixtures of their constituent pure states for each P, as shown in the example for $P = X \otimes X$ above. More concretely, if the number of instances the quantum register collapses to positive eigenspace of $Q = X \otimes X$ for circuits initialized with pure states $|\psi^+\rangle$, $|\phi^+\rangle$, $|\psi^-\rangle$ and $|\phi^-\rangle$ for the case of $P = X \otimes X$ be, p_1, p_2, p_3, p_4 respectively, then, $\text{Tr}\left[X \otimes X\mathcal{G}_\mathbf{j}\rho^{(P)}\right] = \frac{1}{2}\left(\text{Tr}\left[X \otimes X\mathcal{G}_\mathbf{j}|\psi^+\rangle\langle\psi^+|\right] + \text{Tr}\left[X \otimes X\mathcal{G}_\mathbf{j}|\phi^+\rangle\langle\phi^+|\right]\right) = \frac{1}{2}(2p_1 - 1 + 2p_2 - 1)$ and $\text{Tr}\left[X \otimes X\mathcal{G}_\mathbf{j}\hat{\rho}^{(P)}\right] = \frac{1}{2}\left(\text{Tr}\left[X \otimes X\mathcal{G}_\mathbf{j}|\psi^-\rangle\langle\psi^-|\right] + \text{Tr}\left[X \otimes X\mathcal{G}_\mathbf{j}|\phi^-\rangle\langle\phi^-|\right]\right) = \frac{1}{2}(2p_3 - 1 + 2p_4 - 1)$ again, for some choice of two qubit non identity Pauli gate P and sequence $\mathcal{G}_\mathbf{j}$.

After calculating the shifted purity, we further divide by a factor of 4, so as to match with the theoretical calculation of unitarity, which comes out to be one-fourth times the shifted purity calculated empirically, as mentioned in Sec. 2.4. We suspect that authors in [42] didn't mention to divide by 4, to keep the signal strength, i.e., shifted purity values for each iteration high enough for extreme noise channels which have very low unitarity. However, we adopt to divide by a factor of 4 in our implementation. Since, IBM-Q devices are sufficiently robust, the noise induced by the gate operations lie in high unitarity range and hence, we do not gain much by keeping the factor of 4. Additionally, now the shifted purity values itself are good indicators of unitarity of the noise channel composed with itself depth or sequence length many times. However, this is only valid if the SPAM errors are quite low, which we maintain in our simulations.

We also remark here, that following the suggestion of Dirkse *et al.* in [42], we also run the same experiment for given choice of P, Q and $\mathcal{G}_\mathbf{j}$, multiple number of times, to take into account for the within sequence variance in the estimation of unitarity which we call number of samples. *Finally*, we mention here just as a fact that, if $\mathbf{N}_m, \mathbb{M}$ and $\mathbf{S}_m$ be the number of iterations for given depth m, array of depths and number of samples respectively, then for single qubit, number of circuits that needs to be initialized is $18 \times \mathbf{N}_m.\mathbf{S}_m.|\mathbb{M}|$ and for two-qubit it is $900 \times \mathbf{N}_m.\mathbf{S}_m.|\mathbb{M}|$. Here, $|\mathbb{M}|$ is the number of depths or sequence lengths for which the protocol needs to be

implemented. Dirkse *et al.*, give estimates of around 250 for both $\mathbf{N}_m$ and $\mathbf{S}_m$ which amount to around 11 million for single qubit implementation and around 2.25 billion two qubit implementation (for number of depths = 10) different experiment circuits to be executed either on simulator or real device to get unitarity estimate with accuracy error around 10^{-2} and probability of error rate around 10^{-2}. Although, we didn't execute the circuits prescribed many times, but as documented below, we did achieve sub-percent accuracy error rate for sufficiently smaller number of iterations.

In order to test the credibility of our implementation (or rather, proof of satisfactory implementation) of URB single copy protocol, we performed various experiments (both on single and two-qubit registers) considering different noise models, for which we can theoretically calculate the unitarity and tried to benchmark our own implementation. We now present the results of such simulations along with the confidence values. For the purposes of simulating theoretical noise models, we've used Qiskit Aer framework and all simulations have been performed using the *QasmSimulator*.

3.3 URB Results of Custom Noise Simulation

3.3.1 Unitarity of a depolarising noise map

The superoperator $\mathcal{E}$ corresponding to a (completely) depolarising noise over the Hilbert space $\mathbb{C}^{2^n}$, where n is the number of qubits and with depolarising parameter p (treated as the probability, that a given quantum state ρ is unchanged after the noise is applied) is given by:

$$\mathcal{E}(\rho) = p\rho + (1-p)\mathrm{Tr}[\rho]\frac{\mathbb{I}}{2^n}$$

Here, $\mathbb{I}$ is the $2^n \times 2^n$ identity matrix. Using Eq. 2.3 and the fact that $\mathrm{Tr}[\sigma] = 0$, $\forall \sigma \in \mathbb{P}^*$, theoretical unitarity of a depolarising channel is given by,

$$\begin{aligned}
u(\mathcal{E}) &= \frac{1}{4^n - 1} \sum_{\sigma,\tau \in \mathbb{P}^*} \left\{ \mathrm{Tr}\left[p\tau\sigma + (1-p)\mathrm{Tr}[\sigma]\frac{\tau}{2^n} \right] \right\}^2 \\
&= \frac{1}{4^n - 1} \sum_{\sigma,\tau \in \mathbb{P}^*} \left\{ p\mathrm{Tr}[\tau\sigma] \right\}^2 \\
&= \frac{1}{4^n - 1} \sum_{\sigma \in \mathbb{P}^*} p^2 = p^2
\end{aligned}$$

where, we have used the fact that $\sigma, \tau \in \mathbb{P}^*$ are orthonormal with respect to Hilbert-Schmidt inner product and $|\mathbb{P}^*| = 4^n - 1$.
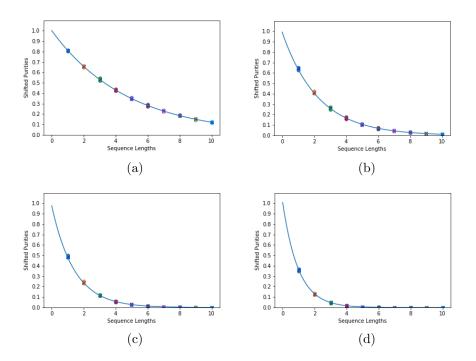
Figure 3.1: URB curves showing the relationship between average shifted purities and sequence lengths (in Clifford unitaries) as a result of simulations of URB single copy protocol for depolarising noise with different values of parameter p, applied to single qubit register. (a) Depolarising parameter $p = 0.9$, estimated unitarity $u = 0.81015$. (b) Depolarising parameter $p = 0.8$, estimated unitarity $u = 0.64081$. (c) Depolarising parameter $p = 0.7$, estimated unitarity $u = 0.49238$. (d) Depolarising parameter $p = 0.6$, estimated unitarity $u = 0.36072$. For each simulation, the sequence lengths were taken to be $m = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$, number of iterations were 15 and for each iteration, the sample size was 5. Variance in the estimated unitarity was found to be 2.3×10^{-5} over 10 iterations of repeating the entire experiment.

Fig. 3.1 shows the exponential curves plotted for single-qubit URB implementation with noisy simulation of (completely) depolarising noise per single-qubit Clifford gate for different values of p. We take the depths or the number of Clifford unitaries applied to be $m \in \{1, 2, \ldots, 10\}$. For each depth, we run 15 iterations and for take 5 samples for given depth and iteration. In each plot, the blue line represents the best fit curve given by $\tilde{q}_m = Bu^{m-1}$, where u is the estimated unitarity from the experiment and m is the number of noisy gates applied and the colored dots are the shifted purities for different iterations and given depth. Concentration of the colored dots along the blue line indicate that the variance in the shifted purities is quite low and is centered around the average shifted purity.

Fig. 3.1 (a), shows the fit curve for $p = 0.9$, which result in unitarity of 0.81015, as compared to theoretical unitarity of 0.81. Fig. 3.1 (b), shows the fit curve for $p = 0.8$, which result in unitarity of 0.64081, as compared to theoretical unitarity of 0.64. Fig. 3.1 (c), shows the fit curve for $p = 0.7$, which result in unitarity of 0.49238, as compared to theoretical unitarity of 0.49. Finally, Fig. 3.1 (d), shows the fit curve

46

for $p = 0.6$, which result in unitarity of 0.36072, as compared to theoretical unitarity of 0.36.

3.3.2 Unitarity of a single qubit bit-flip channel

A mixed unitary noise map is defined as the convex sum of unitary noise channels. Mathematically, it is defined as:

$$\mathcal{E}(\rho) = \sum_{k=0}^{n} p_k U_k \rho U_k^{\dagger}$$

Here, $\{U_1, U_2, U_3..., U_n\}$ is the set of unitary operators and $\{p_k\}_{k=0}^{n}$ is the probability that the quantum state ρ is subjected to unitary noise $U_k \rho U_k^{\dagger}$ associated with U_k.

One particular mixed unitary noise channel which is widely studied, with respect to quantum error correction literature, is the bit-flip channel (for a single qubit) which is given by:

$$\mathcal{E}(\rho) = p\rho + (1-p)X\rho X \tag{3.1}$$

i.e, with probability p, the qubit is left unchanged (applying identity channel) and with probability $1 - p$, Pauli X is applied to the qubit. We are only considered here with single qubit bit-flip channel, though generalisation for multi-qubits is also possible. For multi-qubit case, we have to consider the independence (dependence) of errors in the qubits i.e, do the bit flips occur simultaenously for all the qubits or for any subset of them. This will make the compuation of unitarity a bit more involved and hence, for the sake of theoretically calculating the unitarity, we stick to the simplest case of single qubit bit-flip channel.

Now, we derive the theoretical unitarity of the bit flip channel defined in Eq. (3.1) as

follows.

$$u(\mathcal{E}) = \frac{1}{4^n - 1} \sum_{\sigma, \tau \in \mathbb{P}^*} \left\{ \text{Tr}\Big[\tau(p\sigma + (1-p)X\sigma X)\Big]\right\}^2$$

$$= \frac{1}{4^n - 1} \sum_{\sigma, \tau \in \mathbb{P}^*} \Big(p\text{Tr}[\tau\sigma] + (1-p)\text{Tr}[\tau X\sigma X]\Big)^2$$

$$= \frac{1}{4^n - 1} \sum_{\substack{\tau \in \mathbb{P}^* \\ \sigma = \tilde{X}}} \Big(p\text{Tr}[\tau\tilde{X}] + (1-p)\text{Tr}[\tau X\tilde{X}X]\Big)^2$$

$$+ \frac{1}{4^n - 1} \sum_{\substack{\tau \in \mathbb{P}^* \\ \sigma \neq \tilde{X}}} \Big(p\text{Tr}[\tau\sigma] + (1-p)\text{Tr}[\tau X\sigma X]\Big)^2$$

$$= \frac{1}{4^n - 1} \sum_{\substack{\tau \in \mathbb{P}^* \\ \sigma = \tilde{X}}} (\text{Tr}[\tau\tilde{X}])^2$$

$$+ \frac{1}{4^n - 1} \sum_{\substack{\tau \in \mathbb{P}^* \\ \sigma \neq \tilde{X}}} \Big(p\text{Tr}[\tau\sigma] - (1-p)\text{Tr}[\tau\sigma]\Big)^2$$

$$= \frac{8p^2 - 8p + 3}{3}.$$

Here, $\tilde{X}$ is normalized Pauli X in $\mathbb{P}^*$ (normalized Pauli group for single qubit). We also use the fact that, for a single qubit, $X\omega X = -\omega \;\; \forall \omega \in \mathbb{P}^*$.

The results of our experiments are shown in Fig. 3.2. Similar to the depolarising channel case as before, we take the depths or the number of Clifford unitaries applied to be $m \in \{1, 2, \ldots, 10\}$. For each depth, we run 15 iterations and for take 5 samples for given depth and iteration. In Fig. 3.2(a), we show the fit curve of a bit-flip channel, which does not 'flip' the qubit state with probability $p = 0.975$. The unitarity was found to be 0.935424, while the theoretical unitarity calculated through the derivation above is 0.935. In Fig. 3.2(b), we show the fit curve of the bit-flip channel, with probability parameter $p = 0.95$. The unitarity was found to be 0.876434, while the theoretical unitarity calculated through the derivation above is 0.873. In Fig. 3.2(c), we show the fit curve of the bit-flip channel, with probability parameter $p = 0.9$. The unitarity was found to be 0.772098, while the theoretical unitarity calculated through the derivation above is 0.76. In Fig. 3.2(d), we show the fit curve of the bit-flip channel, with probability parameter $p = 0.8$. The unitarity was found to be 0.624513, while the theoretical unitarity calculated through the derivation above is 0.573.

In the above experiments, we find that in extreme conditions, when the noise strength is considerably high, leads to overestimation of unitarity when URB protocol is used to calculate it. For instance, in the depolarising channel case, the regime of $p < 0.5$
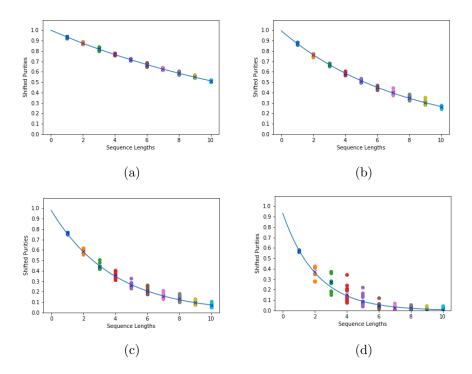
Figure 3.2: URB curves showing the relationship between average shifted purities and sequence lengths (in Clifford unitaries) as a result of simulations of URB single copy protocol for bit-flip channel with different values for probability p, applied to single qubit register. (a) For $p = 0.975$, estimated unitarity wa found to be $u = 0.935424$. (b) For $p = 0.95$, estimated unitarity was found to be $u = 0.876434$. (c) For $p = 0.9$, estimated unitarity was found to be $u = 0.772098$. (d) For $p = 0.8$, estimated unitarity wa found to be $u = 0.624513$. For each simulation, the sequence lengths were taken to be $m = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$, number of iterations were 15 and for each iteration, the sample size was 5. Variance in the estimated unitarity was found to be 3.6×10^{-4} over 10 iterations of repeating the entire experiment.

or in bit-flip channel, if the probability of survival, i.e., the probability that the noise channel will not disturb the state is less than 0.9. This is largely because of the very weak signal strengths, i.e., the shifted purity values for modest increase in depth size. As it is evident from Fig. 3.1 and 3.2, the shifted purities for depth sizes of more than 5, in the 'strong' regimes for respective noise channels, the shifted purities are close to zero. This induces error in the fitting of the exponential curve and as a result, the empirical unitarity is overestimated by the protocol. However, in the high unitarity regimes, i.e., for the cases in which the shifted purities are sufficiently larger than zero, the results of URB protocol are validated by the theoretical prediction with sub-percent variance in the results. We conclude that, our experiments are completely consistent with the theoretical prediction of unitarity and hence, establish the soundness of our implementation.

3.4 Benchmarking noise in native gates of IBM-Q processors

3.4.1 Native Gate URB

In the previous section, we provided evidence for the soundness of our implementation of URB. We considered depolarising channel and mixed unitary channel as toy noise models for gate errors and experimentally calculated the unitarity which is consistent with the theoretical calculations. In this section, we present the main contribution, in which we calculate the unitarity of the noise map generated naturally in the operation of native gates present in two of the quantum processors of IBM-Q devices suite, namely burlington (5-qubit) and melbourne (15-qubit).

For every quantum computing archetype and platform (be it superconducting transmon qubits of IBM or Google or trapped ions archetype manufactured by IonQ), each particular gate operation requested by the user, with respect to running a certain user fabricated quantum circuit, has to ultimately be decomposed into sequence of specific chosen gates (called the native gates) which have dedicated instruction for applying them on the processor, at the time of actual circuit execution on the processor. Since, universal quantum computation is ensured by a universal gate set, for example {CNOT, H, S and T} gates, these native gates are a superset of such a universal gate set (some gates may not belong to the universal gate set but nevertheless, have explicit instruction for it's implementation on the processor). The process of transforming a given circuit, by decomposing each gate operation into sequence of gates from the native gate set is sometimes known as 'compiling' the circuit.

We mention here, that although, URB protocol that has been proposed in the literature is a theoretical procedure which suggests a method to experimentally calculate the unitarity of averaged noise map of a chosen gate set (preferably a unitary 2-design). It is similar to SRB to the extent that both need a gate set and fitting an exponential curve. However, when it comes to implementing the procedure practically, caution must be practiced, regarding the interpretation of it's results. Speaking along the lines of the argument provided by [38] and previously discussed in the Introduction, both the results of SRB and URB are dependent on the actual compilation of the individual clifford gates in the platform/processor. Since, most of the archetypes/platforms do not generally have the whole clifford group in the native gate set, a sub optimal compilation of a clifford gate can affect the results of the experiment. For instance, one may think of implementing identity gate (assuming for the sake of argument, is not a part of native gate set) by applying two CNOTs successively to the target and nearest neighbour qubit. Although, this implementation of identity gate is logically equivalent to an identity operation, a noisy CNOT gate applied twice imparts considerable noise in the form of cross talk (for a single qubit

gate operation) than, a specific native gate for identity operation.

We therefore, because of the reason mentioned in the preceding paragraph, perform a variant of URB protocol (here on referred to as *Native Gate URB*), which is similar in spirit to the proposal of IRB and DRB protocols, that benchmarks the unitarity of the noise of a given native gate from the set of native gate set pertaining to a processor or platform. This variant is although, not drastically different from the proposed URB protocol (here on, referred to as Standard URB or SURB), is mainly different with respect to two facts:
(1) *Firstly*, we no longer require a gate set.
(2) *Secondly*, We do not require sampling of a random sequence from a gate set, since there is no gate set.

The reason lies in the interpretation of the definition of metrics associated with SRB and SURB. The metric of interest for SRB is average gate fidelity and as mentioned in section 2.4, it essentially quantifies the average error rate of a noise channel (think of it, as the average behaviour exhibited by the noise channel averaged with respect to set of pure states). This 'averaging' is signified by the twirling operation of the noise map over the unitary group $\mathbb{U}(2^n)$ and essentially to approximate this operation in practice, we employ the use of a unitary 2-design gate set such as a Clifford group. Hence, the role of a gate set in SRB is important in extracting average gate fidelity of a gate set (as in SRB) or individual gate (as in IRB) by an experimental procedure. On the other hand, as discussed in section 2.4, definition of unitarity is concerned with the unital part of the noise and hence, is directly related to the noise channel. The role of gate set in SURB, is to obtain the average noise map for a gate set (not necessarily a unitary 2-design) and not for a mandatory operation, like twirling, to perform 'averaging' over the pure states. Although Wallman *et al.* initially propose the definition of unitarity in the form of an 'average over pure states', they prove that essentially that reduces to focusing only on the unital part of the channel.

To summarise, unitarity is a concept which discriminates the nature of noise channel and is not a metric about actual behaviour of the noise channel (i.e the error rate of the channel like average gate fidelity or diamond distance) and hence, does not require a mandatory gate set to perform 'averaging' of noise. Hence, there exist separate protocols (DRB, IRB, RRB etc.) which are variants of SRB, but for SURB, all other similar variants are just just 're-interpretation'. In our main contribution, we perform a analog of mix of IRB and DRB protocols (which exist for SRB) for SURB. We discuss our exact procedure used to benchmark unitarity of noise in IBM-Q devices below.

Native Gate URB details: In order to benchmark unitarity of noise map of actual gate operation of native gates in IBM-Q processors we propose the following two

modifications to existing SURB protocol:

(1) There is no gate set that we consider, so no sampling is required.

(2) In each iteration and each depth d of the circuit, single identity gate is applied, (for two-qubit version, we apply $I \otimes I$) followed by the native gate in consideration and this is repeated d times.

The rest of the protocol remains same. We mention here, that although depth of the circuit which played the role of approximating the average noise channel in SURB now, becomes irrelevant for the calculation of unitarity of a fixed native gate, nevertheless, we did perform experiments for various depths by being skeptical about this assumption, simply because the unitarity then, derived from the exponential regression can come out to be more accurate. Again, since sampling of gates is not performed, there is no 'in-sequence variance' and hence number of samples becomes irrelevant, but we still keep the number of samples parameter, which essentially means that we perform same experiment many times for 'averaging' the noise from the native gate. We present compact pseudocode for Native gate URB in Algorithm 3.4.1.

Input: Sequence lengths (depths) $\mathbb{M}$, Number of iterations (or number of random sequences) $\mathbf{N}_m$ for every depth $m \in \mathbb{M}$, Number of samples $\mathbf{S}_m$ for each depth and iteration, native gate to be benchmarked $\mathcal{G}$ and ideal identity gate (on all qubits) $\mathbb{I}$.

Begin:

   For all depth $m \in \mathbb{M}$:

     Repeat $\mathbf{N}_m$ times:

      Compose the sequence $\mathcal{G}_m = \mathcal{G} \circ \mathbb{I}...\mathbb{I} \circ \mathcal{G} \circ \mathbb{I} \circ \mathcal{G}$.

      where, $\mathcal{G} \circ \mathbb{I}$ is applied m times (ignore first $\mathbb{I}$)

      For all non-identity Pauli's P, Q:

       Repeat $\mathbf{S}_m times$:

       Prepare pure states corresponding to,

       $\rho^{(P)} = \frac{I+P}{d}$ and $\hat{\rho}^{(P)} = \frac{I-P}{d}$

       as mentioned in 3.2.1. Apply $\mathcal{G}_m$

       to each state. Measure $E^{(Q)} = Q$ large number

       of times. Estimate shifted purity,

$$q_{\mathrm{s}} = \frac{1}{d^2 - 1} \sum_{\mathrm{P,Q} \neq \mathrm{I}} \{ \mathrm{Tr}[E^{(Q)} \mathcal{G}_{\mathbf{j}} \rho^{(P)}] - \mathrm{Tr}[E^{(Q)} \mathcal{G}_{\mathbf{j}} \hat{\rho}^{(P)}] \}^2$$

      Average over all samples $\tilde{q}_n = \frac{1}{\mathbf{S}_m} \sum_s q_s$

     Average over all iterations $\hat{q}_m = \frac{1}{\mathbf{N}_m} \sum_n \tilde{q}_n$

    Fit exponential curve $\hat{q}_m = Bu^{m-1}$ where, u is

    unitarity and B is a constant absorbing SPAM.

End

Algorithm 3.4.1 Pseudocode for Native gate URB Protocol (single copy implementation).

We mention here, that the unitarity experiments we performed are simulations using *QasmSimulator* of IBM Qiskit, similar to the simulations of custom noise models, discussed in 3.3. We have used the *NoiseModel.from_backend(backend)* method provided by Qiskit Aer, that approximates the noise model of real backend device. We treat this noise as the average noise of the native gate. It simulates the actual noise imparted due to gate operation on qubits, in real backends, by considering an effective depolarising channel followed by thermal relaxation of the qubit. We have considered both of these effects and also considered only the depolarising error (excluding the thermal relaxation of the qubit), but we ultimately perform experiments by not considering the thermal relaxation so that the approximate noise simulated is close approximation of noise induced solely by the gate operation.

3.4.2   Results of Native Gate URB in IBM-Q processors

The native gate set for both burlington (5-qubit) and melbourne (15-qubit) backends consist of the set {CNOT, I, U2, U3} where, I is the identity gate, and U2 and U3 correspond to single qubit rotations parameterized by two and three angles respectively. Fig 3.3 and 3.4 provide the details of the results of the experiments of Native Gate URB. We now elaborate on the actual implementation details of our experiment.

For each processor and native gate chosen, we apply a custom 'unitary' gate which performs the identity operation, followed by the native gate chosen, repeated depth many times in each sample and iteration. We remark here, that this custom unitary gate is *not* decomposed into native gates, since Qiskit allows us to define custom *basis_gates* or gates from which we want our 'compiled' circuit to be executed and we add this unitary gate (essentially, the non-noisy identity gate) and the native gate chosen, to our set of *basis_gates*. This means, in our 'compiled' circuit, the depth of the circuit is composed of the ideal unitary gate performing identity operation and the noisy native gate chosen which will be the sole source of noise in the circuit. The reason we can be sure of ideal identity operation is because no IBM-Q device (real backend) contains 'unitary' gate as it's *basis_gate* and since, we can only simulate the backend noise of native gates of a given processor, therefore, the custom unitary gate we construct is ideal. We also make sure to include this custom unitary gate in our set of *basis_gates* so that it does not decompose into noisy native identity ('id') gate of real backend. We choose to apply ideal identity gate and native gate to be benchmarked in an 'interleaved' fashion because, internally Qiskit tries to optimize the circuit when it compiles the circuit to be run. Since we explicitly want the same noisy gate to be applied many times for each depth experiment, we specifically want Qiskit to *not* optimize the circuit and remove the number of duplicate gates. This case specifically arises for 'cx' gate, for which Qiskit does not apply any 'cx' in the compiled version if, the depth is even and applies a single 'cx' if the depth
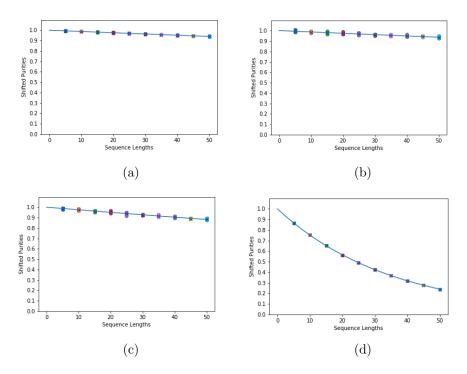
Figure 3.3: Results of experimental implementation of variant of Standard URB procedure, to benchmark unitarity of native gates of burlington processor (display_name: ibmq_burlington). URB curves associated with experiments done with native gates chosen as Identity ('id'), U2 ('u2'), U3 ('u3') and CNOT ('cx') are shown in Fig (a),(b),(c),(d) respectively. The depths used for each benchmarking experiment were, for identity and CNOT gates, $m = 5, 10, 15, 20, 25, 30, 35, 40, 45, 50$ and for U2 and U3 gates, $m = 1, 2, 3, 4$. Number of iterations for each benchmarking experiment $= 15$ and number of samples $= 5$. Unitarity of approximate noise map of 'id', 'u2', 'u3' and 'cx' were found to be 0.998473, 0.998683, 0.997480, 0.971898 respectively.

is odd. By interleaving, an ideal identity gate, we disarm this optimization behaviour.

Fig 3.3 shows the results for the burlington device. Fig 3.3 (a), is the fit curve obtained for the identity gate ('id') of the burlington device. The experiment was performed with the sequence lengths of $m = 5,10,15,20,25,30,35,40,45,50$, number of iterations being 15 and number of samples for the same depth and iteration were taken to be 5. The unitarity was found to be 0.998473, which is high as expected, since identity is a single qubit gate hence, there is little chance of any incoherent error in it's execution. The variance in the shifted purities for each depth value was also found to be low, (which can be seen as colored blobs for different depth values along the curve). This means that, the shifted purity values are concentrated about the mean.

Fig 3.3 (b), is the fit curve obtained for the 'u2' gate which takes two angle parameters. The experiment was performed with the sequence lengths of $m = 5,10,15,20,25,30,35, 40,45,50$, number of iterations being 15 and number of samples for the same depth and iteration were taken to be 5. The unitarity was found to be 0.998683, which is

high as expected, again, since it is a single qubit gate. The variance in the shifted purities for each depth value was also found to be very low.

Fig 3.3 (c), is the fit curve obtained for the 'u3' gate which takes three angle parameters. The experiment was performed with the sequence lengths of $m = 5,10,15,20,25,30,$ 35,40,45,50, number of iterations being 15 and number of samples for the same depth and iteration were taken to be 5. The unitarity was found to be 0.997480, which is again high as expected as it is again, a single qubit gate. The variance in the shifted purities for each depth value was again, found to be very low.

Fig 3.3 (d), is the fit curve obtained for the CNOT ('cx') gate. The experiment was performed with the sequence lengths of $m = 5,10,15,20,25,30,35,40,45,50$, number of iterations being 15 and number of samples for the same depth and iteration were taken to be 5. In this case, the unitarity was found to be 0.971898 which is slightly lower than the single qubit gates and is also expected. Since, the two qubit gates are prone to cross talk between qubits, we suspect that unitarity of CNOT gate must be lower than single qubit gates, as CNOT may induce incoherent errors. However, we find that the unitarity is sufficiently high, signifying that IBM-Q burlington device has significantly less cross talk and as a result, the incoherence in the errors is very low. We argue that, this is expected evidently, due to simple connectivity and less number of qubits in the burlington device.

Moving on to the 15 qubit processor (melbourne), Fig 3.4 (a), is the fit curve obtained for the identity gate ('id') gate of the burlington device. The experiment was performed with the sequence lengths of $m = 5,10,15,20,25,30,35,40,45,50$, number of iterations being 15 and number of samples for the same depth and iteration were taken to be 5. The unitarity was found to be 0.997995. We remark that, this is very slightly lower than the unitarity of identity gate found for burlington device.

Fig 3.4 (b), is the fit curve obtained for the 'u2' gate which is single qubit gate taking two angle parameters. The experiment was performed with the sequence lengths of m = 5,10,15,20,25,30,35,40,45,50, number of iterations being 15 and number of samples for the same depth and iteration were taken to be 5. The unitarity was found to be 0.997861.

Fig 3.4 (c), is the fit curve obtained for the 'u3' gate which is the single qubit gate taking three angle parameters. The experiment was performed with the sequence lengths of $m = 5,10,15,20,25,30,35,40,45,50$, number of iterations being 15 and number of samples for the same depth and iteration were taken to be 5. The unitarity was found to be 0.995872, slightly lower than that of 'u2' gate.

Fig 3.4 (d), is the fit curve obtained for the CNOT ('cx') gate. The experiment was performed with the sequence lengths of $m = 5,10,15,20,25,30,35,40,45,50$, number of iterations being 15 and number of samples for the same depth and iteration were taken to be 5. We found the unitarity to be 0.930389, which is considerably lower compared to the results till now. We therefore conclude that, there is significant cross talk inducing incoherent errors in the case of two-qubit CNOT gate for the 15-qubit
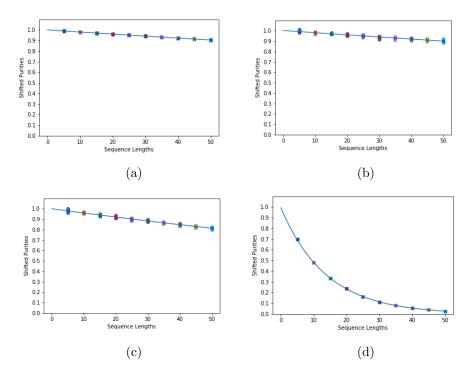
(a) (b)

(c) (d)

Figure 3.4: Results of experimental implementation of variant of Standard URB procedure, to benchmark unitarity of native gates of melbourne processor (display_name: ibmq_16_melbourne). URB curves associated with experiments done with native gates chosen as Identity ('id'), U2 ('u2'), U3 ('u3') and CNOT ('cx') are shown in Fig (a),(b),(c),(d) respectively. The depths used for each benchmarking experiments are, for identity and CNOT gates $m = 5, 10, 15, 20, 25, 30, 35, 40, 45, 50$ and for U2 and U3 gates, $m = 1, 2, 3, 4$. Number of iterations for each benchmarking experiment $= 15$ and number of samples $= 5$. Unitarity of approximate noise map of 'id', 'u2', 'u3' and 'cx' were found to be 0.997995, 0.997861, 0.995872, 0.930389 respectively.

IBM-Q melbourne device although still, it's under reasonable limits. Again, this only strengthens the argument we gave before, that simple connectivity and less number of qubits induce less cross talk among the qubits, while rich connectivity and/or more number of qubits (which is what we see in case of melbourne device) induces more incoherent errors. The results are summarised in Table 3.1.

Table 3.1: Table summarising the Native gate URB experiments on 5-qubit and 15-qubit IBM-Q processors.

Native Gate	Burlington (5-qubit)	Melbourne (15-qubit)
Identity ('id')	0.998473	0.997995
Single-Qubit ('u2')	0.998683	0.997861
Single-Qubit ('u3')	0.997480	0.995872
Two-Qubit ('cx')	0.971898	0.930389

56

3.4.3 Discussion on the Results

Here, we comment on our results and discuss the extent of cross talk in the IBM-Q processors which we have considered in this paper. Since, cross talk errors are correlated [48], cross talk between target qubits (which, the gate operation intended to act on) and neighbouring qubits (extraneous qubits affected by the gate operation) can induce incoherent errors in the target qubits. Conversely, target qubits of a given gate operation can be subjected to incoherent noise after the gate is applied, either due to decoherence of the targets qubits themselves or due to cross talk caused by the gate operation itself. We have already ensured that, the simulation of the backend noise is executed on ideal qubits with no thermal relaxation errors. This rules out the decoherence of the target qubits themselves as the source of 'incoherence' in the noise channel.

We therefore conclude from the results summarised in Table 3.1 that, there is significant cross talk inducing incoherent errors in the case of two-qubit CNOT gate for both 5-qubit and 15-qubit processors. For the rest of the native gates, for both the processors, we conclude that there is little 'incoherence' in the noise and practically the noise remains unitary for low depths. Although, for large circuit depths of more than 30 the 'incoherence' in the noise associated with single qubit gates for both processors become significant, considering they only act on a single qubit.

We also conclude that, the cross talk induced by the CNOT gate for 15-qubit processor induces more 'incoherence' in the noise than the 5-qubit one by observing that the unitarity of CNOT gate in 15-qubit is about 0.04 less than that of 5-qubit processor. We remark here, that we are only able to detect cross talk and it's effect on target qubits in terms of 'incoherence' induced in the noise channel through our Native Gate URB protocol. We cannot comment on the further characterization of the cross talk based on the obtained results. Although, quantifying and characterizing the cross talk based on the concept of unitarity and estimating a figure of merit by employing a randomized benchmarking type protocol might be an interesting question to consider. Nevertheless, the least we can reason is, that simple connectivity and less number of qubits induce less 'incoherence' as a result of cross talk, while rich connectivity and/or more number of qubits (which is what we see in case of 15-qubit device) induces more when applying the same entangling gate such as the CNOT gate. This is however true, only when the 'quality' of qubits in both the processors are comparable i.e., either they are ideal qubits or their decoherence times are of the same order.

Chapter 4

Conclusion and Future Work

In this paper, we experimentally verified the significance of URB, by running a modified protocol that we proposed to benchmark the noise in two of the IBM-Q devices. In these experiments our control over the exact compilation and actual implementation on the IBM-Q devices were to the extent of simply performing Native gate URB successfully. We propose more detailed experiments as future work, in which the user chooses the qubits present on the device (based on the decoherence times of the qubits) as the 'working qubits' and perform URB on these qubits, so that, full characterization of noise can be performed on a given processor, considering it's restricted connectivity. Both SURB and Native Gate URB, require the whole Pauli group to be accessed every time we try to benchmark the coherence in the noise, which scales exponentially with the number of qubits. We propose the modification of Native Gate URB, so that it can be practically tractable also as a future problem. We hope that, this can be of potential use to quickly empirically calculate a figure of merit that can give an idea about the amount of cross talk and incoherence in the quantum computation performed on a given commercial device or archetype.

Bibliography

[1] Moore's Law Wikipedia contributors. Moore's law. Wikipedia, The Free Encyclopedia. June 30, 2020

[2] P. W. Shor SIAM J. Comput., 26(5), 1484–1509.

[3] M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information (10th Anniversary edition) Cambridge University Press, 2010

[4] F. Arute, K. Arya, R. Babbush and *et al.* Nature **574**, 505–510 (2019).

[5] Church-Turing Thesis Wikipedia contributors. Church–Turing thesis. Wikipedia, The Free Encyclopedia. June 20, 2020

[6] P. Rebentrost, M. Mohseni, and S. Lloyd Phys. Rev. Lett. **113**, 130503 (2014).

[7] S. Lloyd, M. Mohseni and P. Rebentrost Nature Phys **10**, 631–633 (2014).

[8] E. Farhi and H. Neven arXiv:1802.06002 [quant-ph] (2018).

[9] E. Farhi, J. Goldstone, S. Gutmann *et al.* Science Vol. **292**, Issue 5516, pp. 472-475 (2001).

[10] P. Rebentrost and S. Lloyd arXiv:1811.03975 [quant-ph] (2018).

[11] H. R. Grimsley, S. E. Economou, E. Barnes *et al.* Nat Commun **10**, 3007 (2019).

[12] S. McArdle, S. Endo, A. A. Guzik, S. C. Benjamin and X. Yuan Rev. Mod. Phys. **92**, 015003 (2020).

[13] D. S. Abrams and S. Lloyd Phys. Rev. Lett. **79**, 2586 (1997).

[14] Nature **543**, 159 (2017).

[15] K. Wright, K. M. Beck, S. Debnath and *et al.* Nat Commun **10**, 5464 (2019).

[16] P. C. Humphreys, W. S. Kolthammer, J. N. M. Barbieri and *et al.* Phys. Rev. Lett. **113**, 130502

[17] J. Preskill Quantum **2**, 79 (2018).

[18] Many Worlds Interpretation of Quantum Mechanics Wikipedia contributors. Many-worlds interpretation. Wikipedia, The Free Encyclopedia. July 5, 2020.

[19] L. Gurvits Proceedings of the Thirty-fifth Annual ACM Symposium on Theory of Computing. p. 10, 2003

[20] C. D. Bruzewicz, J. Chiaverini, R. McConnell and J. M. Sage Applied Physics Reviews **6**, 021314, 2019

[21] M. Kjaergaard, M. E. Schwartz, J. Braumüller and *et al.* Vol. **11**:369-395 , 2019

[22] S. S. Tannu and M. K. Qureshi ASPLOS '19: Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages Pages 987–999 (2019).

[23] Quantum Computing Report: Where Qubits entangle with Commerce

Gottesman-Knill Theorem Wikipedia contributors. Gottesman–Knill theorem. Wikipedia, The Free Encyclopedia. June 12, 2020

[24] B. Schroeder, E. Pinheiro, W. D. Weber Communications of the ACM, Volume 52, Issue 2, 2011

[25] I. L. Chuang, R. Laflamme, P. W. Shor, W. H. Zurek Science Vol. **270**, Issue 5242, pp. 1633-1635 (1995).

[26] E. Farhi, J. Goldstone and S. Gutmann arXiv:1411.4028 [quant-ph] (2014).

[27] E. T. Campbell, B. M. Terhal and C. Vuillot Nature. **549** (7671): 172–179 (2017).

[28] A. W. Cross, L. S. Bishop, S. Sheldon, P. D. Nation, and J. M. Gambetta Phys. Rev. A **100**, 032328 (2019).

[29] Abraham, AduOffei, I. Y. Akhalwaya and G. Aleksandrowicz *et al.* Qiskit: An Open-source Framework for Quantum Computing (2019).

[30] R. Computing Pyquil documentation (2019).

[31] J. Emerson, R. Alicki and K. Życzkowski J. Opt. B: Quantum Semiclass. Opt. **7** S347 (2005).

[32] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland Phys. Rev. A **77**, 012307 (2008).

[33] C. Dankert, R. Cleve, J. Emerson and E. Livine Phys. Rev. A **80**, 012304 (2009).

[34] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, M. B. Ketchen and M. Steffen Phys. Rev. Lett. **109**, 080505 (2012).

[35] A. K. Hashagen, S. T. Flammia, D. Gross and J. J. Wallman Quantum 2, 85 (2018).

[36] R. Harper and S. T. Flammia Phys. Rev. Lett. **122**, 080504 (2019).

[37] W. G. Brown and B. Eastin Phys. Rev. A **97**, 062323 (2018).

[38] T. J. Proctor, A. Carignan-Dugas, K. Rudinger, E. Nielsen, R. Blume-Kohout, and K. Young Phys. Rev. Lett. **123**, 030503 (2019).

[39] J. K. Iverson and J. Preskill arXiv:1912.04319v2 [quant-ph]

[40] Y. R. Sanders, J. J. Wallman and B. C. Sanders New J. Phys. **18** 012002 (2016).

[41] J. Wallman, C. Granade, R. Harper and S. T. Flammia New J. Phys. **17** 113020 (2015).

[42] B. Dirkse, J. Helsen, and S. Wehner Phys. Rev. A **99**, 012315 (2019).

[43] 5-qubit backend: IBM Q team, "IBM Q Burlington backend specification V1.1.4," (2019).

[44] 15-qubit backend: IBM Q team, "IBM Q 16 Melbourne backend specification V2.1.0," (2018).

[45] R. Kueng, D. M. Long, A. C. Doherty, and S. T. Flammia Phys. Rev. Lett. **117**, 170502 (2016).

[46] QuaEC: Quantum Error Correction Analysis in Python

[47] A. W. Cross, L. S. Bishop, J. A. Smolin and J. M. Gambetta arXiv:1707.03429

[48] M. Sarovar, T. Proctor, K. Rudinger and *et al.* arXiv:1908.09855